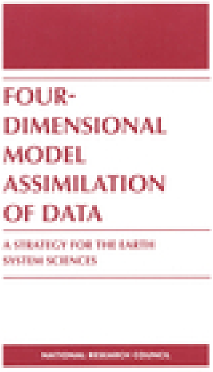
http://www.nap.edu/catalog/1830.html
We ship printed books within 1 business day; personal PDFs are available immediately.



Four-Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences

Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research, National Research Council

ISBN: 0-309-58348-9, 88 pages, 6 x 9, (1991)

This PDF is available from the National Academies Press at: http://www.nap.edu/catalog/1830.html

Visit the National Academies Press online, the authoritative source for all books from the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, and the National Research Council:

- Download hundreds of free books in PDF
- Read thousands of books online for free
- Explore our innovative research tools – try the "Research Dashboard" now!
- Sign up to be notified when new books are published
- Purchase printed books and selected PDF files

Thank you for downloading this PDF. If you have comments, questions or just want more information about the books published by the National Academies Press, you may contact our customer service department toll-free at 888-624-8373, visit us online, or send an email to feedback@nap.edu.

This book plus thousands more are available at http://www.nap.edu.

Copyright © National Academy of Sciences. All rights reserved.
Unless otherwise indicated, all materials in this PDF File are copyrighted by the National Academy of Sciences. Distribution, posting, or copying is strictly prohibited without written permission of the National Academies Press. Request reprint permission for this book.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

Four-Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences

Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research
Board on Atmospheric Sciences and Climate
Commission on Geosciences, Environment, and Resources
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1991

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine.

The National Academy of Sciences is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Frank Press is president of the National Academy of Sciences.

The National Academy of Engineering was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Robert M. White is president of the National Academy of Engineering.

The Institute of Medicine was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Samuel O. Thier is president of the Institute of Medicine.

The National Research Council was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Frank Press and Dr. Robert M. White are chairman and vice chairman, respectively, of the National Research Council.

Support for this project was provided by the National Science Foundation under Grant Number ATM-8819538.
Library of Congress Catalog Card No. 91-61622

International Standard Book Number 0-309-04536-3
Additional copies of this report are available from: National Academy Press 2101 Constitution Avenue, N.W. Washington, DC 20418

S373
Printed in the United States of America

PANEL ON MODEL-ASSIMILATED DATA SETS FOR ATMOSPHERIC AND OCEANIC RESEARCH

DONALD R. JOHNSON, University of Wisconsin, Madison, *Chairman*
JOHN R. BATES, National Aeronautics and Space Administration, Goddard Space Flight Center
GUY P. BRASSEUR, National Center for Atmospheric Research
MICHAEL GHIL, University of California, Los Angeles
ANTHONY HOLLINGSWORTH, European Centre for Medium Range Weather Forecasts
ROY L. JENNE, National Center for Atmospheric Research
KIKURO MIYAKODA, National Oceanic and Atmospheric Administration, Geophysical Fluid Dynamics Laboratory
EUGENE RASMUSSON, University of Maryland, College Park
EDWARD S. SARACHIK, University of Washington
THOMAS T. WARNER, Pennsylvania State University
JOHN S. PERRY, Staff Director
KENNETH H. BERGMAN, Staff Officer
DONALD H. HUNT, Consultant

BOARD ON ATMOSPHERIC SCIENCES AND CLIMATE

JOHN A. DUTTON, Pennsylvania State University, Chairman
JON F. BARTHOLIC, Michigan State University
E. ANN BERMAN, Tri-Space, Inc.
RAFAEL L. BRAS, Massachusetts Institute of Technology
MOUSTAFA T. CHAHINE, California Institute of Technology
ROBERT A. DUCE, University of Rhode Island
THOMAS E. GRAEDEL, AT&T Bell Laboratories
DAVID D. HOUGHTON, University of Wisconsin, Madison
EUGENIA KALNAY, National Oceanic and Atmospheric Administration
RICHARD S. LINDZEN, Massachusetts Institute of Technology
SYUKURO MANABE, National Oceanic and Atmospheric Administration
GERALD R. NORTH, Texas A&M University
JAMES J. O'BRIEN, Florida State University
JOANNE SIMPSON, National Aeronautics and Space Administration

Ex-Officio Members

ERIC J. BARRON, Pennsylvania State University
PETER V. HOBBS, University of Washington
CHARLES E. KOLB, Aerodyne Research, Inc.
DONALD J. WILLIAMS, The Johns Hopkins University
JOHN S. PERRY, Staff Director
WILLIAM A. SPRIGG, Associate Staff Director
KENNETH H. BERGMAN, Staff Officer
DONALD H. HUNT, Consultant
FRED D. WHITE, Consultant

COMMISSION ON GEOSCIENCES, ENVIRONMENT, AND RESOURCES

M. GORDON WOLMAN, The Johns Hopkins University, *Chairman*
ROBERT C. BEARDSLEY, Woods Hole Oceanographic Institution
B. CLARK BURCHFIEL, Massachusetts Institute of Technology
RALPH J. CICERONE, University of California, Irvine
PETER S. EAGLESON, Massachusetts Institute of Technology
HELEN M. INGRAM, University of Arizona
GENE E. LIKENS, New York Botanical Garden
SYUKURO MANABE, National Oceanic and Atmospheric Administration
JACK E. OLIVER, Cornell University
PHILIP A. PALMER, E. I. du Pont de Nemours & Company
FRANK L. PARKER, Vanderbilt University
DUNCAN T. PATTEN, Arizona State University
MAXINE L. SAVITZ, Allied Signal Aerospace Company
LARRY L. SMARR, University of Illinois, Urbana-Champaign
STEVEN M. STANLEY, Case Western Reserve University
KARL K. TUREKIAN, Yale University
SIR CRISPIN TICKELL, Radcliffe Observatory
IRVIN L. WHITE, N.Y. State Energy Research and Development Authority
JAMES H. ZUMBERGE, University of Southern California
STEPHEN RATTIEN, Executive Director
STEPHEN D. PARKER, Associate Executive Director
JANICE E. MEHLER, Assistant Executive Director
JEANETTE SPOON, Administrative Associate
CARLITA PERRY, Administrative Assistant

Contents

Four-Dimensional Model Assimilation of Data: A Strategy for the Earth System Sciences

Executive Summary

The purpose of this report is (1) to define the need for a nationally focused program to generate routinely research-quality, model-assimilated geophysical data sets and (2) to outline a basic strategy to implement an integrated national geophysical archive system to ensure the availability of such data sets to serve a broad range of national needs.

The report concisely surveys the current applications and usefulness of research and operational model-assimilated atmospheric and oceanic data sets, assesses current activities for the development of model assimilation technology and new applications, and identifies the pressing national need in the 1990s—to manage and utilize effectively the overwhelming volume of earth system data already scheduled from greatly enhanced ground-based and satellite observing systems. A basic strategy for the earth system sciences is outlined, namely, to use the power and consistency of proven model assimilation technology to generate routinely assimilated geophysical data sets for both operational and long-term use and to maintain these data sets in an integrated national geophysical archive system designed to ensure the ready availability of these data sets to the scientific and public policy communities. By extending this technology backward in time to about the beginning of the development of model-based geophysical data in the 1950s through reanalyses with state-of-the-art model assimilation, the consistency and integrity of these data sets will provide information and value (e.g., for the study of climate and global change) that significantly exceed those of the incoming observations.

Geophysical data assimilation is a quantitative, objective method to infer the state of the earth-atmosphere-ocean system from heterogeneous, irregularly distributed, and temporally inconsistent observational data with differing accuracies. The current method for atmospheric and oceanic sciences uses numerical prediction models, which incorporate our best knowledge of the physics and dynamics of the continually evolving earth system, to integrate these observations into temporally, spatially, and internally consistent data sets that provide a better description over time of the system's state than that provided by the raw observations. These model-assimilated data sets have proven to be extremely valuable as initial conditions for numerical weather prediction and also for a great variety of diagnostic research studies that have increased our understanding of atmospheric and oceanic behavior and, more generally, of how the climate system works. The increased understanding has, in turn, led to the development of improved numerical models and data assimilation techniques.

Currently, several major operational forecast centers produce model-assimilated data sets for the atmosphere as part of their numerical prediction activities. Use of these global-scale data sets as initial conditions for numerical weather prediction models has resulted in significant increases in forecast accuracy. New assimilation techniques, such as sequential estimation and variational data assimilation, along with the installation of more powerful computer systems, will lead to further increases in accuracy. Data assimilation with developing regional and mesoscale models, and future extension into stratospheric models, will lead to advances in our understanding of and predictive capability for atmospheric circulations on these scales as well.

The state of ocean modeling and the production of model-assimilated data sets are less advanced than for the atmosphere, mainly because the amount of oceanic data collected routinely is much less than for the atmosphere and because there traditionally has been no requirement for real-time ocean forecasting. The relatively slow circulation of the oceans only partially compensates for the paucity of observations. Nevertheless, pilot studies on the feasibility of producing useful model-assimilated data sets have been carried out and look promising. Model-assimilated data sets are recognized as the best means of ensuring that the various oceanic observations are as dynamically consistent with one another as possible, just as in the atmosphere.

Another area where model-assimilated data sets have the potential to combine large amounts of heterogeneous observational data into a coherent whole is in the analysis of the hydrological cycle. This cycle consists of three dominant subsystems: the oceanic source, the atmospheric conveyor, and the land surface catchment and runoff. Each of these presents formidable, but not insurmountable, challenges in observational data handling and integration into model representation of the hydrological cycle. More detailed and more extensive observational coverage of evaporation, precipitation, change in soil moisture, and other parameters, including ocean and land surface processes, that define the hydrological cycle is needed.

An important goal of the U.S. Global Change Research Program is to develop coupled interactive models that link the atmosphere, oceans, land masses, and biosphere into a comprehensive whole. Model-assimilated data sets produced by such coupled model systems would provide an improved description of the entire global system and of all its interacting components. Atmospheric chemistry and oceanic biogeochemical modeling would also be included in this comprehensive modeling of global processes. Data assimilation techniques have been used successfully with limited chemical transport models for both the troposphere and the stratosphere. It should be emphasized that further progress toward development of comprehensive global models that produce model-assimilated data sets useful for global monitoring will require the implementation of new observing systems, many of which are already scheduled for the 1990s.

Research problems and studies on a wide range of spatial and temporal scales will benefit from the availability of high-quality model-assimilated data sets. In the atmosphere these include global, regional, and mesoscale weather systems; planetary waves; low-frequency variability; atmosphere-ocean and atmosphere-land interactions; dynamics-chemistry-radiation interactions; and the whole hydrological cycle. Similar research studies in the oceans will also benefit as observational coverage is increased to provide global oceanic data sets. Pilot studies indicate that interactively coupled atmospheric and oceanic models lead to more realistic modeled behavior of both fluids.

A major challenge of the coming decades is in tracking global and regional climatic trends. Model-assimilated data sets currently in existence were produced mainly for operational forecasting purposes and are not temporally and internally consistent over long time periods because of frequent changes in the early model development period in numerical models and assimilation techniques and different practices with regard to incorporation of off-schedule and research data. Therefore, a model data assimilation analysis of the entire useful climatic record over the past 40 years is essential for removing the impacts of any model biases on the assimilated data sets and producing a temporally and spatially consistent data set for the study of climatic trends. Although such an extended analysis of past data would be a relatively expensive undertaking, the need for tracking and predicting global change in terms of scientific and economic factors justifies the effort.

Data assimilation includes a systematic, structured, and open-ended learning process. Learning occurs through iterative confrontations of different types of "current" observations from nature with each other and with the predicted model state, which serves as the "background' field. This confrontation provides a disciplined and rich opportunity for learning about the behavior of the geophysical systems, the quality of the observations, the interpretation of observational evidence, and the accuracy of the assimilating model. The background fields, for example, are sufficiently accurate to be useful for detecting observations with uncharacteristically large errors. However, instances have occurred in which the model rejected observations that truly represented a significant change, resulting in a subsequent incorrect forecast. Methods have been developed to catch these critical events, but it cannot be said that state-of-the-art models have entirely resolved this problem.

Data assimilation has proven especially valuable in isolating the systematic error characteristics and biases of satellite-based remote sensors and also observational stations and should therefore have a major role in both assessing the quality of future satellite and ground-based data and assuring their use to best advantage. Furthermore, since the error characteristics of model-assimilated data sets can be estimated, the data sets themselves are also used routinely to validate the performance of prediction models.

A survey of existing archives of model-assimilated data sets indicates that they are frequently incomplete, are saved in a variety of locations and in a variety of formats, and are not easily accessible in general to the research community. Nevertheless, they provide good starting points for development of an integrated national archival system that would have uniform formats and provide ease of access to users.

Future changes in technology and cost will facilitate the process of compact archiving of vast data sets and will make their use by individual researchers much more feasible. New data storage technology, including publishable media such as compact data (CD) disks, will permit scientists to store significant amounts of data at their own computer workstations. New devices and computer software routines for imaging data fields will also facilitate use of model-assimilated data sets by researchers.

Large increases in heterogeneous observational data from greatly enhanced ground-based and satellite observing systems in the future will require an expanded use of model assimilation of data and the production of model-assimilated data sets, both for operational and research purposes. For example, in the United States the ongoing National Weather Service (NWS) Modernization Program, involving implementation of new observing systems like the WSR-88D Doppler radar, profilers, advanced satellite systems, and the Automated Surface Observing System (ASOS), is expected to result in an overall 100-fold increase in the amount of observational data that must be integrated into a dynamically consistent data set for operational prediction and warning purposes. Similar challenges are raised in physical and biogeochemical oceanography and for the earth system by the increased amount of data expected in the future. Research on data assimilation systems capable of managing the new observing systems of the 1990s is presently under way, but this effort will need additional support to be operational in time for the new observing systems already scheduled. The operational centers and the larger research community must contribute jointly to the design of these systems to ensure that the needs of the broad community are met.

Continuing advances in computer technology and modes of parallel computer operation will permit more sophisticated data assimilation techniques to be used at national forecast centers and laboratories, will allow oceanic and other geophysical model-assimilated data sets to be routinely generated, and will increase the feasibility of reanalyses of long data sets from the past as the state-of-the-art technology requires. Affordable facilities for acquiring, processing, and evaluating large data sets, along with appropriate software, are expected to become increasingly available for use by individual researchers.

CONCLUSIONS

The panel reached four major conclusions:

1. Four-dimensional (space and time) data assimilation as a subdiscipline of geophysical sciences is fundamental for the synthesis of diverse, temporally inconsistent, and spatially incomplete observations into a coherent representation of an evolving geophysical system.
2. Viewed as an integral element of the scientific process, four-dimensional model assimilation of geophysical data is a systematic, quantitative, objective, iterative means of inference and testing aimed at advancing understanding and prediction of nonlinear dynamical geophysical systems where interactions occur continually among relevant physical, chemical, and biogeochemical processes.
3. Since the physical and dynamical consistency of model-assimilated data sets results in a level of information and added value that significantly exceeds that of the incoming observations, assimilation data sets have been and will be used even more extensively by the scientific community for diagnostic, predictive, and process studies, supplemented by original observations when needed.
4. Atmospheric model data assimilation is a proven strategy in an advanced state of development. Although still in an early stage of development for the earth sciences as a whole, model data assimilation is strategically situated to address the pressing national need to observe and understand global change as it occurs in the coming decades. The immediate need is to

develop and provide a nationally focused capability to synthesize, test, and utilize for understanding and prediction the information from the greatly enhanced new ground-based and satellite observing systems already scheduled for the next two decades.

RECOMMENDATIONS

The panel recommends the following strategy for a nationally focused program for four-dimensional model assimilation of data for the earth system:

1. The strategy for the nationally focused program will:

- Develop and expand applications of model-based data assimilation efforts to interdisciplinary areas that are necessary for integrated earth ocean-atmosphere-biogeochemical models.
- Implement an integrated, multicenter, nationally focused archive system for model-assimilated and-tested data sets and provide for ready access to these sets by the scientific community over the long term.
- Provide for continuing scientific exchange and collaboration among the various groups and individuals engaged in geophysical data assimilation, particularly scientists involved with the Earth Observing System Data and Information System; the national meteorological, oceanographic, and climate centers; and the High Performance Computer and Communications Initiative.
- Provide for the generation of routine research-quality, model-assimilated and-tested geophysical data sets to serve a broad range of national endeavors, including climate and global change research and predictions.
- Establish a working group to develop an implementation plan for this nationally focused program.

The strategy and implementation plan should include the following specific actions:

2. To validate and maintain quality control of new types of remotely sensed and experimental in situ geophysical data, including research data from field programs, the application of operational data assimilation models is essential. Funding agencies should routinely provide sufficient funds and computer capacity for this purpose, including provision for timely communication of such new and experimental data sets to designated operational assimilation centers.
3. A coordinated national program should be implemented and funded to develop consistent, long-term assimilated data sets (extended back to about

1950) for study of climate and global change. This effort will reanalyze, with a state-of-the-art global-and regional-scale data assimilation model, all atmospheric and oceanic data available since about 1950 in order to produce the best possible, validated, temporally and spatially consistent data sets for the study of climate and global change. Model biases should be identified and eliminated as far as possible at this time. This reanalysis should be repeated as advances in the state of the art require.

4. Data assimilation models for the mesoscale should be developed in concert with regional prediction models. The systems should integrate data from the enhanced observational capabilities of the coming decade that will be provided through the NWS Modernization Program, advanced satellite systems, and the Earth Observing System (EOS). These models should be capable of integrating specially observed data with the conventional data stream and also provide for realistic responses to various types of system forcing for the time and space scales emphasized. The mesoscale assimilation systems should also ensure effective nesting with larger-scale analysis systems.
5. In order to provide ready access by the scientific community, a multicenter geophysical data archive system, electronically linked for maximum effectiveness for assimilated data set management, transfer, and usage, should be created.

- The data management activity for the multicenter archiving system will ensure routine compilation and availability of observed, model-assimilated, and model-predicted data sets in a structural format jointly adopted and coordinated with the Earth Observing System Data and Information System (EOSDIS). There is a vital need to archive together with the data sets the modeling codes, information on the input data and how they were processed, and so forth for the scientist to be able to evaluate and understand the archived data sets.
- Accessibility to archived model-assimilated data sets should be part of a preplanned, service-oriented archive system, including on-line electronic links, low-cost publication media for use on individual workstations, routine inclusion of metadata and software for unpacking and manipulating data sets, and high-quality imaging capabilities.

6. Interdisciplinary and discipline graduate education and research opportunities in four-dimensional geophysical data assimilation should be created. To provide essential expertise to cope with the 100-fold increases in observational data from greatly enhanced observing systems in the 1990s and beyond, the development of graduate education, including advanced degree and research opportunities in the fundamental subdiscipline of four-dimensional data assimilation, is needed. Immediate support should be

provided for graduate courses and research fellowships in university departments of atmospheric, oceanic, global change, and related earth system sciences. These programs should be explicitly coordinated with related system development, such as EOSDIS, the High Performance Computer and Communications Initiative, and national computer-linked networks.

1

Introduction

The complexity of the earth-ocean-atmosphere system and the nonlinear processes governing its dynamics pose formidable challenges to our scientific understanding and to our capabilities to extend the range of useful prediction. The governing dynamical equations for this nonlinear geophysical fluid system are evolutionary, rather than equilibrium, equations. For the foreseeable future, observations of this system will be irregularly distributed in space and time, will be obtained using a great variety of instruments and methods, and will continue to be inhomogeneous in accuracy. In addition, the observations are frequently too sparsely distributed to resolve adequately the complex spectrum of scales and processes involved. These limitations and the need to describe the state of the evolving geophysical system as accurately as possible have led to the development of geophysical model data assimilation as a rational method to infer the state of the system within which dynamical, physical, chemical, and biological processes can be coupled interactively.

With funding from several federal agencies, the National Research Council's Board on Atmospheric Sciences and Climate established the Panel on Model-Assimilated Data Sets for Atmospheric and Oceanic Research. This study is the report of the panel. The panel's charge was to prepare a concise report that defines a nationally focused effort to routinely generate model-assimilated, research-quality data sets for the needs of the 1990s, to briefly survey the current usefulness of such data sets as a starting point for application to the full range of anticipated enhanced geophysical data streams, and to outline a basic strategy for a national archive system that in the future will effectively serve a broad range of research and operational programs.

CONCEPT OF GEOPHYSICAL MODEL DATA ASSIMILATION

The basis for geophysical model data assimilation may be described as the four-dimensional representation of a unified, dynamically evolving geophysical system by a mathematical model. This model has the capability to predict the dynamic changes occurring in the system, accept the insertion of new observational data distributed heterogeneously in time and space, and blend earlier information and current information objectively under rigorous quality control. This data assimilation technology leads to an estimate of the state of the system that is more complete and more accurate, and thus of higher value, than can be achieved from direct analysis of a single set of observations taken at a particular time. The model is used to mathematically extrapolate information in time in accordance with basic physical and chemical laws, moving from the "recent past" to the "present state" of the system. This predicted present model state constitutes the "background" field against which a current set of new observations are subjected to quality control, are interpreted, and are synthesized with the extrapolations of all previous information. In other words, the background field, gridded or spectral, is adjusted using the assimilated information from the current set of observations, thus providing the "updated" model state. In this process the current observations are filtered and interpolated in space and time in a manner that is as consistent as possible with the dynamical model and the presumed true state of the system. Starting from the updated state, the model is then integrated into the future, to provide a forecast and, concurrently, a background field for the ingestion of the next set of observed data, and so the process continues, whereby the state of the system and its evolution are continually described.

The process of data assimilation provides the most complete and accurate synthesis of our theoretical and a priori knowledge with all available observational information, and the result is the best estimate of the state of the system and its evolution that can be obtained with all the available data. The resulting synthesis is called a model-assimilated data set (MADS), consisting of gridpoint values or spectral coefficients for the specified time.

For data assimilation to be successful, it is essential that the assimilating model have useful predictive skill in providing a mechanism for filtering and interpolating the observations. If the model has no predictive skill in the extrapolation of earlier information forward in time, the impact of the earlier information cannot be expressed. The ability of a skillful model to transport information from data-rich areas into data-sparse areas is particularly valuable, although even a very accurate model cannot substitute completely for bad or incomplete data.

ASSESSMENT OF DATA ASSIMILATION IN ATMOSPHERIC SCIENCES

Successful prediction for a dynamical system depends critically on accurate knowledge of the state of a system. In this regard, data assimilation has proven to be a powerful method to infer the current state of the atmosphere. The approach has been exploited in operational meteorology with unusual success and has contributed to remarkable gains in forecast skill over the last decade (Figure 1).

Dynamically consistent fields of model-assimilated data have also proven

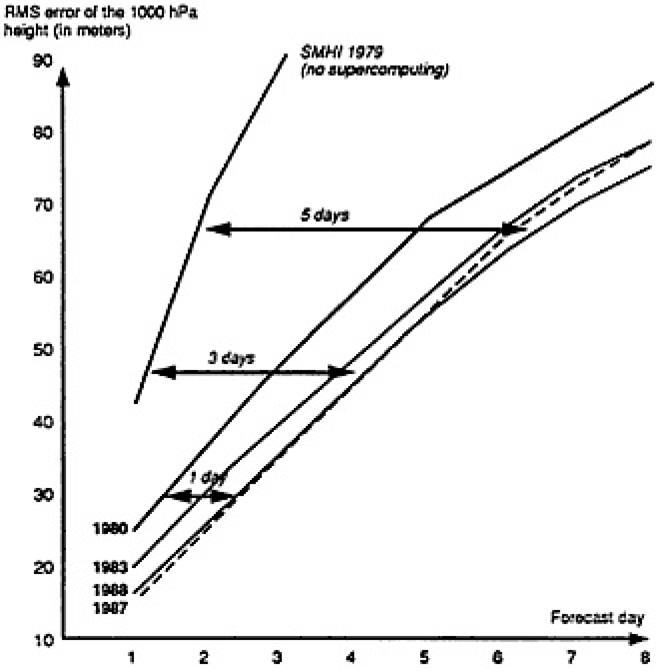


FIGURE 1 RMS error of the 1000 hPa height as a function of the forecast length for a quasigeostrophic model in 1979 (Swedish Meteorological-Hydrological Institute) and a primitive equation model in 1980, 1983, 1987, and 1988 (European Centre for Medium Range Weather Forecasts) during the first quarter (winter) of the year (Lange, 1988).

to be extremely valuable for scientific studies of the circulation within the earth-atmosphere system, in terms of both the system's structure and the processes that maintain the circulation. Assimilation models are unique in their ability to produce fields of complex nonlinear processes with a physical and dynamical consistency that is unattainable by any other method of analysis. As value-added sources of information, these model-assimilated data sets constitute a priceless resource for diagnostic and predictive studies of atmospheric and oceanic circulation ranging from mesoscales (as small as 2 km) through planetary scales (as large as 10,000 km).

DATA ASSIMILATION VIEWED AS PART OF A SYSTEMATIC LEARNING PROCESS

The data assimilation process is one of continual confrontation between theoretical and observational knowledge, as expressed by the current observations from nature and by the predicted model state, which is a function of the previous observations from nature. This confrontation presents a rich opportunity for a structured, iterative, and open-ended learning process about the behavior of atmospheric, oceanic, and other geophysical systems; the quality of the observations; the interpretation of observational evidence; and the accuracy of the assimilating model.

In the last decade, observational research has used model-assimilated data sets as a primary source material for the study of a wide range of atmospheric phenomena and processes. There are many reasons for this development. Model-assimilated data sets constitute an internally consistent time series of global three-dimensional fields whose accuracy can be estimated. Since the gridded fields are regular, the data are in a very convenient form for scientific analysis—that is, dynamical and physical processes involving differentiated and integrated quantities can be readily calculated. Because of the internal consistency, important but in many cases inadequately observed dynamical and physical processes (e.g., atmospheric mass divergence) can be estimated, and theories relating to these processes can be developed and tested. Model-assimilated data sets have also been validated in many studies through the use of independent observational data that were not utilized in the assimilation process. In such studies the validation frequently is a two-way process, where the independent data are used to test the quality of the model-assimilated data sets, and the data sets in turn prove useful for interpretation of the observations. Thus, model-assimilated data sets have contributed substantially to both our theoretical and diagnostic understanding of the atmosphere.

The knowledge gained from theoretical and diagnostic studies, noted above, is the first of four important aspects of the learning process associated with the development and application of data assimilation. The second aspect, and one of the most striking results to date, has been the development of powerful methods to monitor data and to identify important systematic errors in the measurements of many different in situ and remote meteorological observing systems. For example, the ability to identify operationally even modest errors at the most isolated observing locations has furnished dramatic proof of the power of data assimilation to synthesize a predictive model's knowledge of atmospheric behavior with available data into a coherent and accurate picture of that behavior.

The third significant element in the learning process from data assimilation has been the increased effectiveness in the use of remotely sensed data. For example, experiences from the Global Weather Experiment (GWE) (also called the First GARP Global Experiment [FGGE]) revealed that data assimilation with research models benefited substantially from the evaluation and quality control of remotely sensed observational data streams by operational centers. Hence, for the full potential of remote sensing to be developed in the future, it is essential that data from new systems be provided in a timely fashion to operational centers capable of assimilating the data. For example, studies of the quality of wind and temperature retrievals from remotely sensed data, through comparisons with the background fields produced in data assimilation, have exposed many serious limitations in current retrieval procedures for remotely sensed data. Many of these difficulties arise because the measured quantity, usually a radiance, is an integrated nonlinear function of the variable being inferred, such as a vertical profile of atmospheric temperature. Many shortcomings in the current procedures can be removed through variational procedures that explicitly recognize the integrated and nonlinear nature of the remotely sensed data. These procedures reduce the difficulties arising from the integrated nature of the measurements by using information from the background field to compensate for the nonlinearity of the radiative transfer equation. This new approach to the use of remotely sensed data exemplifies how experience with data assimilation generates an impetus for more accurate methods of assimilated data interpretation.

A fourth and vital aspect of the learning process associated with data assimilation is the stimulus provided for the improvement of models used in assimilation and prediction. Such models are by no means perfect. Systematic studies of the errors of the background forecast in the assimilation process, and of longer-range forecasts, have been a powerful stimulus to the detection of defects in models. Missing processes have been identified, known processes have been specified more accurately, and specifications of poorly represented interactions between different processes have been advanced. As assimilation and interpretation of the available observations are improved and the range of observations used in the process is widened, the learning processes inherent in data assimilation become more and more effective. The success of this learning process indicates that data assimilation is the best means known to gain a comprehensive and internally consistent synthesis of all available geophysical data.

CURRENT STATUS OF DATA ASSIMILATION

Currently, several major operational forecast centers use model-based data assimilation to produce four-dimensional (space and time) analyses of atmospheric circulation as a routine starting point for numerical weather prediction. Remarkable advances in both global data assimilation and numerical weather prediction models have been made in the past 15 years. A doubling of the period of time over which a given level of forecast accuracy is maintained has occurred through advances in data assimilation that were largely stimulated by the GWE research and development. Similar data assimilation procedures are applicable to oceanic data and are expected to be applied, for example, in the Tropical Ocean and Global Atmosphere (TOGA) program and the World Ocean Circulation Experiment (WOCE).

Despite the importance of model-based data assimilation for operational prediction, only limited resources have been specifically focused on the development of model-assimilated data sets for research, with the exception of those provided for the GWE and for a few isolated case studies. Currently, the assimilated data sets being made available for research are generated as a by-product of operational weather prediction. Not all operationally generated model-assimilated data sets are readily accessible in general by the larger scientific research community in the United States or abroad. The present practices for archiving and accessing observed and model-assimilated data are inadequate to meet the future needs of the scientific community, particularly for data sets that include consistently all available geophysical data.

APPLICABILITY OF DATA ASSIMILATION TO THE EARTH SYSTEM SCIENCES

The need is emerging for research-quality model-assimilated data sets across a broad spectrum of interactive endeavors known as the earth system sciences (Hollingsworth, 1989). Significant endeavors are beginning in climate and global change, long-range weather prediction, mesoscale weather, hydrology, atmospheric chemistry, physical and biogeochemical oceanography, and land surface processes. Crucial to our understanding of the earth-ocean-atmosphere system is knowledge of the present state and the continuing changes of fundamental properties and trace constituents. Thus, a need exists for a nationally focused effort during the coming decades to routinely provide research-quality assimilated data sets and to ensure their ready accessibility to the larger scientific community. Through a national focus, priorities for scientific exchange among the several communities engaged in data assimilation endeavors must be established, whereby the full potential from technological advances in global and regional observational systems and computational capabilities will be realized. There is an immediate need to assimilate all available atmospheric and oceanic data with a state-of-the-art assimilation model, so as to produce a best-to-date interpretation of the available data record, particularly for study of climate and global change. Such analysis of long records from the past will require the marshaling of financial and manpower resources at the national and even international levels.

The immediate goals of this report are to review the current status of data assimilation and the application of model-assimilated data sets for both operational prediction and scientific research. The panel's recommendations are aimed at ensuring the availability of assimilated data sets for broad national needs in the coming decades. As a starting point for action, the report emphasizes the need for an integrated national effort for the generation, archiving, and service-oriented publication of model-assimilated data sets that will serve a broad range of operational and research programs in atmospheric, oceanographic, and earth sciences. To ensure that the needs of the coming decades are met, the panel includes in this report a recommendation that an integrated national program be developed to provide the focus for and implementation of the full range of effort needed to meet these needs.

2

Data Assimilation Development

INTRODUCTION

The purpose of model assimilation of data from various observing systems, with an irregular distribution in space and time, and with varying and incompletely known error properties, is to produce a four-dimensional depiction of the atmosphere and oceans, with regular distribution of field variables with known error properties in space and time. For such a "movie" depiction, the numerical model used for assimilation must demonstrate a physical consistency with the geophysical system being described in that the associated errors should be as small and well known as possible. The degree to which this goal is achieved depends, for a given set of observing systems, on the data assimilation methodology used.

The existing methodology for atmospheric data and models is briefly reviewed in the next section. It gives reasonably satisfactory results for the mass field and the horizontal winds but leaves much to be desired as to the vertical velocities and the humidity field. In oceanography, data assimilation is now being explored aggressively, but no satisfactory global model-assimilated data set (MADS) of the mass and horizontal velocity fields has emerged as yet (Haidvogel and Robinson, 1989).

Scientific requirements are formulated in this report for improved atmospheric data assimilation; for oceanographic model-assimilated data sets of a quality at least comparable to present atmospheric model-assimilated data sets; and for complementary model-assimilated data sets on surface hydrology, cryospheric components of the climate system, and the relevant chemical and biogeochemical components. These scientific requirements clearly suggest the need for continued development of improved and highly flexible data assimilation methods.

PRINCIPLES AND METHODS

In many current operational systems the data are grouped into convenient (6-hour) groups, and a data assimilation process is carried out, using three modules: analysis, initialization, and forecast. The forecast model is usually a state-of-the-art numerical weather prediction (NWP) model. The forecast field from the earlier data provides the background field (or first guess) for the analysis of the new data. A statistical approach is widely used in the analysis module in current operational systems and is known variously as optimum interpolation (OI) or statistical interpolation (Gandin, 1963; McPherson et al., 1979; Lorenc, 1981).

Operational implementation of the OI approach requires resolution of a number of practical issues. It is difficult to invert a matrix corresponding to a global data set; up to now, a series of local calculations that involve a number of compromises on data selection, continuity between adjacent analysis volumes, multivariate relationships, and so on has been done. Three-dimensional methods to eliminate these compromises are in an advanced state of development. Four-dimensional assimilation systems based on the Kalman filter and on variational analysis that are currently under development will relax limitations and exploit a combination of dynamics and statistics more fully.

An important aspect of the OI approach is that it is a multivariate algorithm that can exploit linear multivariate relations between different variables. The final analysis is a linear combination of the "structure functions" of the forecast error correlations. Thus, any linear constraint imposed on these structure functions is satisfied within its domain of validity by the analyzed fields. Current operational systems impose constraints on the analysis increments, which provide for approximate hydrostatic balance everywhere and approximate nondivergent balance in the extratropics. There is ample empirical justification for these constraints.

The OI algorithm can only use data that are linearly related to the model variables. Data that are nonlinearly related to the model variables (e.g., satellite radiances) must be transformed to variables such as temperature or humidity before use in OI by a retrieval process. The transformation process can introduce many errors. For this reason there is considerable interest in developing more general methods that can use data that are nonlinearly related to model variables.

Quality control of observational data is a critical consideration for accurate analysis because of the unstable nature of atmospheric predictions contaminated by erroneous observations. Operational centers use modified OI routines to perform multivariate cross-checks of each datum against all other data in the vicinity as well as against the forecast field used as a background field.

The meteorological "primitive" equations describe traveling atmospheric disturbances with two inherent time scales: (1) a fast time scale (phase speed ~300 m/s) associated with gravity wave motions and (2) a slow meteorological time scale (phase speed ~50 m/s) associated with large-scale Rossby wave disturbances. An initialization module ensures a smooth start to a forecast by controlling the amplitude of rapidly propagating gravity wave "noise," which is introduced in the course of each analysis step. This control on the "noise" is critical for the quality control procedures employed at the next analysis time since it ensures that the next background field is reasonably accurate and noise free. In the future the fully compressible equations will be employed for data assimilation involving nonhydrostatic phenomena.

CONTINUOUS DATA INSERTION

In the past, the National Aeronautics and Space Administration (NASA) pioneered the first continuous data insertion method (Charney et al., 1969; Ghil et al., 1979). Two groups are currently engaged in the operation and development of a continuous data insertion scheme using a general circulation model (GCM) —the United Kingdom Meteorological Office (UKMO) and the Geophysical Fluid Dynamics Laboratory (GFDL) of the National Oceanic and Atmospheric Administration (NOAA). In the future it is envisaged that several different types of continuous methods may emerge.

In the continuous data assimilation of the U.K. system, the observed data are repeatedly inserted at each time step of the forward integration, using the relaxation technique for injection of data (Lorenc, 1976; Lyne et al., 1982). Observations are first interpolated vertically to the model's level, then horizontally to model gridpoints, and the increments of observed values are used to correct the model values. For the assimilation the Newtonian nudging technique is adopted (Davies and Turner, 1977; Hoke and Anthes, 1976).

The recent GFDL system (Stern et al., 1985; Stem and Ploshay, 1991) is an extension of one employed for the original First GARP Global Experiment (FGGE) analysis at the GFDL (Miyakoda et al., 1976), but it has been modified considerably. Model-assimilated data sets are produced by a GCM, which may be either a spectral or grid model. The observational data are grouped beforehand into 2-hour windows.

The values at the insertion points of model grids (Gaussian grid in the case of a spectral GCM) are found by interpolating the data increments at observation points three-dimensionally with OI. Univariate, rather than multivariate, OI is used because the GCM is expected to provide internal consistency among variables in the continuous scheme.

The insertion data consist of the current solution of the assimilation model (unaffected by the initialization), plus slower gravity modes and all Rossby modes associated with OI-determined incremental changes to the background values. The GCM assimilates these data immediately and produces a continuous stream of model-assimilated data sets.

RESEARCH AND DEVELOPMENT IN DATA ASSIMILATION: THE KALMAN FILTERAND ADJOINT METHODS

At present, the two most promising areas of research and development in data assimilation methods are the extended, nonlinear *Kalman filter* and the *adjoint method*, each with a number of possible simplifications and variations. Research on the Kalman filter was pioneered in the United States (Kalman, 1960; Kalman and Bucy, 1961; Jarwinski, 1970; Gelb, 1974; Bierman, 1977; Bucy and Joseph, 1987). Meteorological applications (Ghil et al., 1981; Parrish and Cohn, 1985; Ghil, 1989) and oceanographic applications (Budgell, 1986; Miller, 1986; Bennett and Budgell, 1987; Miller and Cane, 1989) are being carried out at the Institute of Ocean Sciences of the University of British Columbia, New York University, Oregon State University, Scripps Institution of Oceanography, the University of California at Los Angeles, the University of Rhode Island, and a number of Soviet research institutes. Operational implementation is contemplated at NASA's Goddard Laboratory for Atmospheres.

The adjoint method was pioneered in the Soviet Union and France (Marchuk, 1974; Penenko and Obraztsov, 1976; Le Dimet and Talagrand, 1986). Metéeorological applications (Lewis and Derber, 1985; Talagrand and Courtier, 1987; Derber, 1989) and oceanographic applications (Bennett and McIntosh, 1982) are being pursued at the GFDL, the Laboratoire de Météorologie Dynamique in Paris, the Massachusetts Institute of Technology, NOAA's Miami laboratories, and the University of Oklahoma, among others. Operational implementation is being considered by the European Centre for Medium Range Weather Forecasts (ECMWF) and the National Meteorological Center (NMC).

In principle, both methods try to minimize the distance in phase space between a system trajectory, constrained by model dynamics, and the existing data over a given time interval (Ghil and Malanotte-Rizzoli, 1991). In the adjoint method the constraint is "strong," that is, the model is supposed to be nearly exact; in the Kalman filter approach model, errors are explicitly incorporated. This explicit modeling of errors, while desirable in principle, imposes an additional computational burden. On the other hand, the Kalman filter only needs to handle data at a given instant in time rather than over the entire time interval of interest, as does the adjoint method.

Both methods require linearization of the variational problem they attempt to solve. For the extended Kalman filter this linearization occurs at successive moments in time about a given state of the system. For the adjoint method it occurs in phase space about a given trajectory. Combinations of the two approaches are possible; one might use the Kalman filter only at large intervals (e.g., every week or month) in order to compute the covariance matrices of observational and model errors. The inverses of these matrices can then be used as relative weights for current observations and model background fields in the adjoint method. It should be noted that the adjoint method uses future as well as past data and is therefore well suited for delayed-mode (as opposed to real-time) applications. A Kalman "smoother" also uses future as well as past data.

For operational applications of one or both of these techniques in the future, additional computing power will be needed at the major operational centers. This need is part of the nationally coordinated program presented in the president's fiscal year 1992 budget to Congress in a report by the Committee on Physical, Mathematical, and Engineering Sciences of the Federal Coordinating Council for Science, Engineering, and Technology of the Office of Science and Technology Policy: *Grand Challenges: High Performance Computing and Communications*. As the report's summary states: "The HPCC program is driven by the recognition that unprecedented computational power and capability [are] needed to investigate and understand a wide range of scientific and engineering `grand challenge' problems. These are problems whose solution is critical to national needs. Progress toward solution of these problems is essential to fulfilling many of the missions of the participating agencies. Examples of grand challenges addressed include: prediction of weather, climate and global change. . . ."

3

Current and Future Applications of Model Assimilation Systems

In this chapter current applications of model-assimilated data sets in the areas of global atmospheric circulations, mesoscale atmospheric circulations, physical oceanography, the global hydrological cycle, and atmospheric chemistry are reviewed. Development is quite advanced in some of these areas but is only beginning in others, so that the range of achievement to date is quite wide. The best developed data assimilation theory and practice are found in the area of assimilation for global atmospheric circulations and are discussed first. The importance of future applications for the Earth Observing System (EOS) and other programs is briefly discussed.

GLOBAL ATMOSPHERIC CIRCULATION

As discussed in the previous chapter, current operational and research data assimilation systems at the larger global circulation assimilation centers fall into two main families, intermittent insertion and continuous insertion systems. Intermittent insertion is used in the meteorological centers of Canada; France; Japan; Australia; in the United States at the National Oceanic and Atmospheric Administration's (NOAA) National Meteorological Center (NMC), the U.S. Navy's Fleet Numerical Oceanography Center (FNOC), and the National Aeronautics and Space Administration's (NASA) Goddard Laboratory for Atmospheres (GLA); and at the European Centre for Medium Range Weather Forecasts (ECMWF). The continuous insertion system is used at the United Kingdom Meteorological Office (UKMO) and the Princeton Geophysical Fluid Dynamics Laboratory of NOAA. Within atmospheric science, both methods have been under development for 15 years or more and are now mature applications in the atmospheric sciences. The theories of optimal interpolation and normal-mode initialization have been extensively developed and applied in both strands of development.

These developments have resulted in important increases in both analysis and forecast accuracy in midlatitudes. Understanding of the problems of tropical dynamics and physics has also advanced, with consequent improvements in tropical data assimilation. For example, operational global models now show levels of skill for tropical cyclone track forecasts that are comparable with the best specialized models for track forecasts.

Important new developments in data assimilation methods are likely to be implemented at operational centers over the next 5 years. These developments are expected to offer more accurate interpretations of remotely sensed data and a more consistent description of the time evolution of the atmosphere. When used along with new observing technologies in global assimilation systems of very high resolution (50 km), these methods will offer the prospect of substantial improvement in accuracy for both analysis and prediction.

For purposes of climate assimilation efforts, the most pressing defects of current model-assimilated data sets arise mainly because of insufficient and relatively poor accuracy of observations of the divergent wind and humidity fields, particularly in the tropics. Because of these defects, descriptions in assimilated data of these fields are largely determined by the predictive component of the assimilation model from other fields (nondivergent wind, temperature, pressure) and so are dependent on the assimilation system used to make the inference.

A second limitation of the operationally produced data sets is the difficulty encountered in studies of the climatic variability of parameters that are sensitive to details of the assimilation systems. Because operational systems have been undergoing constant development, studies of phenomena such as the interannual variability of diabatic heating (based on operational data sets) may have superimposed a nonphysical component of variability caused by changes in the assimilation system itself, in the observing network, or in other nonphysical components. During the proposed reanalysis of operationally produced data sets, the state-of-the-art assimilation system employed can be used to identify these sources of variability.

Further limitations on data sets are imposed by the truncation limit of the assimilation system and by the fact that quality control decisions can never be perfect. An example illustrates the relation between these issues: Suppose a ship in a Norwegian fjord reports a strong local wind. The wind may bear little direct relation to the synoptic-scale flow describable on an assimilation grid of 200 km. A data assimilation system may correctly reject the (good) observation, because the observation describes a phenomenon below the resolution of the analysis. Problems of quality control and resolution are always encountered in the vicinity of intense phenomena (e.g., the core of a hurricane) that are below the resolution of the assimilation system.

Quality control algorithms apply many checks to observations: Position checks, internal consistency checks, time consistency checks, climatological checks, checks against adjacent data, and checks against the background field. Quality control algorithms are tuned empirically to maximize the likelihood of accepting good, and rejecting corrupt, observations. Because of the uncertainties inherent in such a probabilistic exercise, bad observations are sometimes accepted and good observations are sometimes rejected. The consequences of erroneous decisions for the resulting forecast can occasionally be serious. Nevertheless, the overriding difficulty in data-sparse regions is paucity of data rather than quality control of data (Meteorological Office, 1987).

MESOSCALE[1] ATMOSPHERIC CIRCULATIONS

Limited-area operational models in the United States, Europe, and Japan that utilize data assimilation systems routinely produce background states for weather prediction. With some exceptions, these background states are determined by intermittent updating of the meteorological fields (Hoke et al., 1989; Golding, 1989), even though plans generally call for next-generation systems to employ continuous assimilation procedures. The three-dimensional variable fields produced are, at best, only marginally adequate for the spatial resolution of many mesoscale meteorological processes within the domain of the limited-area models. The finest-scale operational model available in the United States—the Regional Analysis and Forecast System (RAFS) of the National Weather Service (NWS)—has a grid increment of 80 km on its highest-resolution "C" grid, which spans the North American continent (Hoke et al., 1989). Therefore, the RAFS will resolve best those phenomena with wavelengths longer than about ~800 km. Its suitability for production of assimilated data sets is limited to the synoptic scale and the higher end of the meso-alpha scale, which ranges from 200 to 2000 km. The RAFS analysis does not satisfy the needs of scientists who wish to

---

[1] Mesoscale processes are those that occur on scales of 2 to 2000 km, such as thunderstorms, tornadoes, local heavy rain and snowstorms, flash floods, windstorms, and downslope winds. This scale is subdivided into meso-alpha scale (200 to 2000 km) and meso-beta scale (2 to 200 km).

study phenomena with scales smaller than ~800 km. Thus, model-assimilated data sets that resolve all meso-alpha and meso-beta scales of atmospheric phenomena need to be created by special-purpose research models, as do data sets for geographic regions not covered by operational limited-area models. Such models are currently being developed, for example, at the NOAA Forecast Systems Laboratory (FSL) and the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, and at several universities (Colorado State University, Pennsylvania State University, University of Oklahoma, University of Wisconsin at Madison, Florida State University, University of Illinois, and Drexel University).

Research models generally employ physical process parameterizations that are 2 to 3 years more advanced than operational models. These parameterizations can be specifically designed for particular applications. Research models with variable horizontal and vertical resolutions are generally relocatable (not tied to a specific geographical grid) and can be optimized for the phenomenon being studied. They can be used to generate research-quality data sets on all desired scales. On the other hand, the use of limited-area *operational* models to produce data sets has the advantage that special modeling efforts are not required, with the result that a much larger number of meteorological cases can be made available for research study.

In many studies of mesoscale physical processes, model-based data sets have been generated without the assimilation of any synoptic-scale data during the forecast cycle (except for lateral boundary conditions) because the use of such data through reinitialization would result in the loss of the mesoscale structures generated by the simulation. This is feasible because of the importance of lateral boundary information as determined from observations reflecting the interaction between the mesoscale and larger scales and the relatively good forecasts of atmospheric structure and processes that are often possible for 12-to 72-hour time periods.

The implementation of new technologies in data acquisition, such as the atmospheric profiler and the WSR-88D Doppler radar networks in the NWS Modernization Program, provides the opportunity for the development of four-dimensional data assimilation methods on the mesoscale. For example, NOAA's FSL has begun the development of a mesoscale assimilation model, the Mesoscale Analysis and Prediction System (MAPS). Other examples involving community efforts are the Penn State/NCAR mesoscale model and Colorado State University's Local Analysis and Prediction System (LAPS), which assimilates Doppler radar, satellite soundings, and mesoscale surface observations every hour in real time. In the area of mesoclimatology, the Penn State/NCAR mesoscale model has been used with a four-dimensional data assimilation routine that employs Newtonian relaxation for production of about a 1-year mesoclimatology of flow over North America for the Environmental Protection Agency (EPA) (Seaman. 1989).

PHYSICAL OCEANOGRAPHY

The state of numerical modeling and model assimilation in physical oceanography is far less advanced than that of its atmospheric counterparts for a simple reason: the amount of ocean data collected routinely is orders of magnitude less than that collected for the atmosphere. This is primarily due to the fact that society has not required oceanic forecasts to the same extent that it has weather forecasts. The net result is that the global oceans are poorly sampled (especially at depth) and consequently are poorly modeled and poorly understood.

The lack of data is partially compensated for by the slowness of most ocean changes. If, as was often assumed until recently, the ocean were in a steady equilibrium at depths below the level of seasonal variability, then all data, no matter when collected, would be useful in defining this idealized mean state. However, in the last two decades, relatively rapid (i.e., time scale of months) disruptions in oceanic circulation in near-equatorial regions and slower changes (decades) at intermediate depth in high-latitude regions have been detected. Changes at greater depth on still longer time scales (hundreds and thousands of years) are inferred from the paleoclimatic record and predicted by coupled atmosphere-ocean simulations of green-house warming.

At the present time, for practical reasons, data assimilation experiments in the ocean are concentrated on the shorter time scales of weeks to years. Interest and activity in the field have been growing rapidly. An overview of the subject is provided by a special double issue of *Dynamics of Atmospheres and Oceans* (Vol. 13, Nos. 3–4).

There are currently three distinct types of activities in the area of oceanographic data assimilation and production of model-assimilated data sets. The first is an oceanographic mesoscale prediction activity (Robinson and Leslie, 1985; Robinson et al., 1986, 1989). This activity involves predicting the state of the ocean over a relatively small (100 x 100 km), open region using boundary data as continuous input. In a certain perspective this prediction converts data along the boundary into model-assimilated data sets throughout the region by means of a dynamically consistent interpolation in both space and time using the equations of motion.

The second activity first simulates "data" by running ocean model controls and then assimilates this simulated data into model runs starting from different initial conditions (Philander et al., 1987; Moore et al., 1987; Anderson and Moore, 1989). The use of simulated, rather than actual, data allows for studies of model-based data assimilation in situations where real data are either nonexistent or too sparse to be useful. While the results of these studies are generally optimistic compared to studies using real data, some very useful general principles can be obtained. For example, methods of data assimilation can be examined (Long and Thacker, 1989a,b; Miller and Cane, 1989), the amount and types of data needed for useful initialization of models can be estimated (Philander et al., 1987), and fixes for various technological problems can be tested (Moore, 1989).

As an example of the usefulness of this simulated data assimilation, Philander et al. (1987) found that a few meridional sections of upper-ocean thermal data alone are adequate to initialize the equatorial upper ocean. Under these circumstances the equatorial undercurrent is generated to the correct amplitude by the model without the use of direct-current observations. On the other hand, Anderson and Moore (1989) showed that both oceanic thermal and current data are necessary to accurately initialize an equatorial Kelvin wave.

It should be pointed out that these studies assume that the simulated data are perfect. It is not at all obvious that actual measured thermal data, say with expendable bathythermograph (XBT) data, would yield the same results.

The third type of activity is the assimilation of actual data into numerical ocean models to produce model-assimilated data sets. Leetmaa and Ji (1989) have used an experimental ocean circulation model to assimilate XBT observations taken in the tropical Pacific and have produced monthly fields of currents and thermal structure. The assimilation of data has successfully corrected model-induced diffusive drifts in the vertical thermal structure and has created model-assimilated data fields downstream of where actual data were introduced. The results suffer from a basic problem characteristic of this type of large-scale data assimilation, namely, that the system is forced by imperfectly known winds, resulting in an oceanic response that is sometimes incompatible with the observed data. The monthly data assimilations therefore sometimes oscillate between states inferred by previous upstream data and states forced by new in situ observations. These problems did not arise, however, in the work of Miller and Cane (1989), who used a simpler model of the tropical Pacific but a more sophisticated assimilation method.

A similar assimilation of real XBT data has been performed for the Atlantic (Morlire et al., 1989) for a 1-year period, with data being introduced gradually throughout each month. The results showed a similar improvement in the model thermal structure and a notable improvement in currents for the simulated year. While improvements were significant, discrepancies still occurred between the observed and modeled data, presumably due to inaccuracies in the imposed surface forcing.

Another example is the work of Derber and Rosati (1989); they used a global ocean general circulation model and a continuous injection technique for the assimilation. The results using the Comprehensive Ocean-Atmosphere Data Set (COADS) and the Master Ocean Observations Data Set

(MOODS) over a 13-month period are encouraging, although further improvements are required.

The dilemma in assimilating data in the presence of imperfect surface forcing can be treated in two separate ways. The first is to make the winds part of the fields that can be corrected by the internal ocean data. Thus, through mutual adjustment, accurate ocean data can be used to correct both the ocean and the wind fields and to provide the most consistent combination of forcing and response fields; this is the approach of Thacker and Long (1988). The second approach, not yet tried, would be to assimilate both atmospheric and oceanic data into a coupled atmosphere-ocean model. The dynamics of such a model would adjust the surface fluxes internally, and, in principle, an optimal state of the coupled atmosphere-ocean system should emerge.

As a final note, not a single oceanographic field program as yet has taken the value-added approach of presenting the wide diversity of field observations in the form of model-assimilated data set output of a numerical ocean model, even though such an output is the best means of ensuring that the various and diverse observations are as dynamically consistent with each other as possible.

HYDROLOGICAL CYCLE

The hydrological and energy cycles of the earth system are intimately linked through transports and radiative effects of water vapor in the atmosphere and by transformations between liquid, solid, and vapor states. Rather than dealing with the entire system, the disciplines of meteorology, hydrology, and oceanography have traditionally focused on separate subsystems of the global hydrological cycle.

Atmospheric Subsystem

In the development of the data assimilation/analysis process of atmospheric forecast models, the accurate specification of initial dynamical and thermal fields has, until recently, received higher priority than the specification of hydrological fields. However, the hydrological fields are now receiving increased attention as forecast models become more sophisticated (Tiedtke et al., 1988).

Atmospheric hydrology is the "fast component" of the global hydrological cycle and plays a fundamental role in weather processes and prediction. Consequently, the atmospheric subsystem is the only component of the hydrological cycle for which global model-assimilated data sets are currently being generated as part of operational numerical weather prediction.

When viewed from an atmospheric perspective, the thermal and hydrological conditions at the interface with the ocean and land surface constitute atmospheric boundary conditions. For weather forecasts up to a few days in advance, it is important to have a reasonably accurate description of the initial boundary conditions. However, evolution of the boundary conditions is typically slow relative to the time scales of day-to-day synoptic variability; therefore, accurate incorporation of time-varying boundary conditions in the models has up to now received less attention in operational short-range weather prediction. In contrast, changing boundary conditions are crucial for resolving seasonal and interannual climate variability (Shukla, 1985). As integrated earth system models are developed under the impetus of the climate and global change programs, development of data assimilation will need to address the entire hydrological cycle (see Large-Scale Field Experiments and Biospheric Studies section).

Land Surface Subsystem

There are important differences in the nature of the current operational models and the character of assimilated data sets between meteorology and hydrology. First, there is a fundamental difference in the most suitable coordinate systems for description and prediction. It is convenient to specify global atmospheric fields on a regular grid or in terms of spectral components. In contrast, the natural "grid element" for surface hydrological operations (e.g., river forecasting and soil moisture accounting computations) is the irregularly shaped drainage basin defined with respect to a stream gauging station at the outflow point.

Second, land surface hydrology has traditionally been focused on the water balance of drainage basins whose area is small relative to synoptic spatial scales. However, it should be noted that this spatial-scale mismatch continues to narrow as the resolution of operational atmospheric prediction models increases. The mismatch is reflected in the sampling by the surface-based observations transmitted by the Global Telecommunications System (GTS) of the World Weather Watch (WWW). The WWW surface-based system was designed primarily to resolve synoptic-scale variability. Even over land, where meteorological observations may be relatively dense, a mismatch remains between the synoptic-scale sampling and the much smaller scales of hydrological variability associated with terrain, diurnal variability, planetary boundary layer processes, and the intermittent subsynoptic character of precipitation systems.

The focus of surface hydrology on individual drainage basins, together with the treatment of precipitation as a measured or specified input to the system, has muted any requirement for global data sets. Thus, no system of international data exchange comparable to that for the atmosphere has been developed for traditional surface hydrological data (e.g., high-resolution precipitation data and model-generated evapotranspiration and soil moisture estimates). Consequently, a great deal of surface hydrological data exist for many regions that are not generally made available to the operational meteorological centers or the scientific community for inclusion in model-assimilated data sets.

Ocean Surface Subsystem

The ocean represents the inexhaustible reservoir for the global hydrological cycle. Its primary linkage with the other components is through surface exchanges, that is, precipitation versus evaporation and inflow of fresh water from the land margins. Resultant variations in surface salinity affect the stratification and vertical exchange processes and contribute to large-scale current systems. Along with ice cover, they play an important role in determining the surface heat fluxes. In addition, the hydrological cycle over the ocean plays a very important role in determining the buoyancy flux at higher latitudes and hence determines to a certain extent the intensity of the thermohaline circulation of the ocean.

GLOBAL PRECIPITATION

The measurement deficiencies of the hydrological cycle are severe. Many parameters have never been measured except on a local experimental basis; they have to be inferred from empirical relationships or dynamical balance considerations. New satellite-based observations, combined with more effective use of data from existing satellite systems and improvements in the data assimilation and analysis process of global forecast models, offer the only realistic prospects for significant improvement in the description of the global hydrological cycle.

Precipitation, perhaps the most fundamental hydrological parameter, serves as an important example. Precipitation is adequately measured only over a few well-instrumented land areas of the earth. Furthermore, the transmission of daily values is frequently haphazard, and their use in computing monthly averages is questionable at best. Generally, one must rely on the delayed transmission of monthly averages. Precipitation is also an output of present-day forecast models, but the surface observational coverage is inadequate in most regions for verifying the forecast values.

Satellite-based observations, coupled with an adequate "ground truth" program for calibration and analog development, provide the only realistic prospect of ultimately obtaining global precipitation measurements. There is optimism that the means now exist for significant improvement in the estimation from satellite data of precipitation for periods as short as a few days, particularly in the tropics.

No single space-based instrument is currently capable of providing all the information needed to generate precipitation estimates. The possible measurement techniques include using visible and infrared (IR) imagery together to infer precipitation from cloud height and brightness, scattering-mode passive microwave techniques to infer precipitation intensity from the scattering of ice particles, absorption-mode passive microwave techniques to infer precipitation intensity from the brightness temperature of a rain column of known depth, and active microwave techniques to infer precipitation intensity and condensation profiles from scattering and attenuation from raindrops.

The ability to use observations from the established network of geostationary satellites to obtain space-and time-averaged precipitation estimates over the global tropics has led to the establishment of the Global Precipitation Climatology Program (GPCP), sponsored by the World Climate Research Program (WCRP), principally in support of the Tropical Ocean and Global Atmosphere (TOGA) program (World Climate Research Program, 1986; Arkin and Ardanuy, 1989). The GPCP now produces estimates of 5-day precipitation averaged over areas of 2.51° latitude by 2.5° longitude for the tropics and subtropics using an IR thresholding technique. The project is planned to continue for the 10-year period 1987–1996.

LARGE-SCALE FIELD EXPERIMENTS AND BIOSPHERIC STUDIES

There is growing recognition of the importance of viewing the atmosphere, oceans, land masses, and biosphere as interacting components of a comprehensive earth system (National Research Council, 1990). If the complete system is to be understood, each component must be observed and modeled in the context of its impacts on and control by the other components.

Global studies of the physical climate system and the biogeochemical cycles will require the combination of large-scale simulation models, operating in both the predictive and assimilation modes, and a wide range of data types, ranging from standard meteorological observations to specialized ground-truth efforts for validation of the satellite algorithms.

In the area of land-surface-atmosphere interactions, there are a number of significant problems that must be overcome before global modeling studies can be conducted in a realistic, quantitative way (P. Sellers, University of Maryland, personal communication, 1989). Two major problems are:

1. Model design: Most land surface models currently in use in general circulation models (GCMs) or other large-scale models are based on one-dimensional models of the soil-plant-atmosphere system developed by biologists and agronomists to understand local plant-scale controls on radiation, interception, evapotranspiration, and photosynthesis. For the most

part these models have been used without change to describe area-averaged processes operating over scales of several hundred kilometers. Little work has been done to check the validity of this usage or to develop more sophisticated spatial integration techniques.

2. Initialization and validation: Even if "perfect" biosphere-atmosphere models existed, they would be of little use in the absence of effective methods of initializing them and validating their performance. Clearly, satellite remote sensing offers the greatest promise for providing global fields of surface type, photosynthetic capacity of vegetation, albedo, and soil moisture, but the links between the satellite observations and these biophysical quantities either have not been developed or have not been tested with convincing rigor.

These two problems have been and will continue to be addressed by a series of large-scale field experiments. For example, the following experiments are some that have taken place during recent years:

1. Hydrological-Atmospheric Pilot Experiment (HAPEX) (Andre et al., 1986). This 1986 experiment was conducted in southwestern France over a 100-km^2 area. Surface and airborne heat flux measurements were combined with meteorological and hydrological observations to validate a mesoscale modeling effort. The experiment provided data that allowed an evaluation of the contribution of different land cover types to the regional heat flux field and the resultant influence on local circulation.
2. First International Satellite Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE) (Sellers et al., 1988). FIFE is an intensive effort designed to specifically address the modeling scale issue and the remote-sensing inversion task. The field phase of the experiment was executed from 1987 to 1989 over a 15-km^2 area in Kansas that was heavily instrumented with automatic meteorological stations and surface flux rigs. To date, analysis of the data has shown that there are strong relationships between satellite observations in the short-wave part of the spectrum and the biophysical parameters of photosynthetic capacity and canopy water content. An effort is being made to combine all the field observations with a one-dimensional surface-atmosphere column model to calculate surface states and fluxes.
3. Another French-led experiment, HAPEX-II/Sahel, is planned for execution in 1992 in Niger. It will combine elements of the FIFE and HAPEX experiments.

ATMOSPHERIC CHEMISTRY

Concentrations of trace gases in the atmosphere have not been systematically measured on a global scale over long periods of time. Thus, very limited data assimilation has been performed by the atmospheric chemistry community. An important exception is the case of ozone and related species, such as chlorofluorocarbons and methane, and a few other species such as water vapor, nitric acid vapor, and nitrogen dioxide, which have been measured by instruments such as the Total Ozone Mapping Spectrometer (TOMS), the Solar Backscatter Ultraviolet (SBUV) spectrometer, the Limb Infrared Monitor of the Stratosphere (LIMS), and the Stratospheric and Mesospheric Sounder (SAMS) on board the Nimbus 7 satellite, as well as by the Solar Mesosphere Explorer. These data have been used to verify the consistency of the presently accepted chemical scheme for the middle atmosphere and to derive the distribution of fast-reacting radicals produced from chemical transformations of the measured trace gases.

Studies of specific chemical events in the troposphere over limited geographical domains, usually in connection with observational campaigns, are performed by coupling a rather detailed chemical kinetic code with a transport model. The winds used to simulate transport during these events require model-assimilated data that are provided by mesoscale meteorological models. Difficulties in these simulations include accounting for possible rapid vertical transport and chemical conversion in convectively active clouds and correctly specifing the conditions at the boundaries of the limited domain involved. This type of approach is currently used in connection with acid rain studies and pollution episodes, such as those during which high concentrations of tropospheric ozone are observed over continental areas.

Data assimilation techniques have also been used in a limited number of cases to understand the transport and chemical transformations of trace gases in the stratosphere. Assimilation algorithms have, for example, been used to reproduce the behavior of nitric acid and ozone during a stratospheric warming event. Again, in this type of study, only the dynamical fields (temperature, geopotential height, and winds) area assimilated and used to predict transport of the species.

Finally, a limited number of attempts to assimilate the chemical data themselves have been made by replacing, in a chemical transport model, the calculated distribution of a chemically active trace species by the corresponding observed distributions. An estimation of the effect of this substitution on the calculated concentrations of other species that are present provides a test to validate the chemical scheme adopted in the model.

Data assimilation using models with predictive capability for chemical species will probably be used more extensively when planned future satellites, such as the Upper Atmosphere Research Satellite (UARS) and the polar platforms of the Earth Observing System, begin to provide continuous observations of trace constituents on the global scale over a period of several years.

EARTH OBSERVING SYSTEM

The Earth Observing System (EOS) is a program being planned by NASA in collaboration with the European Space Agency (ESA) and the National Space Development Agency (NASDA) of Japan to utilize satellite measurements to study climatic and global changes of the earth system. The program calls for the deployment of a large variety of earth-sensing instruments on platforms in low-earth polar orbit. Plans for EOS also include a number of earth-observing instruments as attached payloads on the planned low-declination NASA Space Station. The full system is scheduled to operate for a 15-year period, presumably long enough to obtain a useful time series of earth observations for monitoring global change.

From the point of view of producing global atmospheric data sets for climate research, a global satellite observational capability is essential. Two of the most important EOS instruments will be NASA's Atmospheric Infrared Sounder (AIRS) and the Laser Atmospheric Wind Sounder (LAWS). AIRS will measure radiances in more than 115 spectral bands with the goal of obtaining temperature retrievals of 1°C accuracy and 1-km vertical resolution throughout the depth of the troposphere. It will be a prototype for the next-generation operational IR sounder. LAWS is a Doppler lidar system for active tropospheric wind measurement in regions without opaque clouds. The expectation is that an accuracy of 1 to 5 m/s with a 1-km vertical resolution and an average horizontal spacing of 100 km can be achieved.

NASA is planning an EOS Data and Information System (EOSDIS), which will provide for the reception, processing, storage, and distribution of EOS data. An important component of EOSDIS will be a four-dimensional data assimilation system, which will use both EOS and conventional data in delayed mode to produce research-quality data sets for the earth-atmosphere-ocean-land system. The Global Modeling and Simulation Branch of the Goddard Laboratory for Atmospheres (GLA), with cooperating teams from the National Meteorological Center (NMC) and the Australian Bureau of Meteorology Research Center, has been selected by NASA to develop a high-resolution four-dimensional atmosphere-ocean-land data assimilation system for EOS. The panel's recommendation for a nationally focused geophysical data assimilation program calls for close coordination with the EOSDIS program.

USE OF MODEL-ASSIMILATED DATA SETS FOR RESEARCH

Data sets from assimilation of observations and from simulations employing assimilated data as initial state information are widely used by the meteorological community. For example. more than 200 institutes in over

50 countries on six continents have acquired model-assimilated data sets from ECMWF. Most data requests are for research purposes. The heaviest users are in the developed countries, but there are also many requirements for the data sets in the developing world. The advantage of these data sets is that they systematically combine information from many different satellite and ground-based sources, a task that is beyond the resources of small university or national research groups.

The use of model-assimilated data sets for the study of large-scale atmospheric flow was pioneered in the middle to late 1970s by Wallace and Lau at the University of Washington and by Blackmon at the National Center for Atmospheric Research. Their studies of the diree-dimensional structure of midlatitude atmospheric circulation led to a considerable advance in our understanding of low-frequency variability in the atmosphere.

The Global Weather Experiment (GWE) provided, for the first time, a detailed global view of atmospheric structure. The availability of global data accelerated the development of techniques of global data assimilation. The resulting model-assimilated data sets provided good quality descriptions of the mass and wind fields and formed the basis for a wide range of research studies. Vital information on large-scale tropical phenomena such as El Niño and the 30- to 60-day oscillations has been determined through the use of these data sets. In recent years model-assimilated fields have also been invaluable for study of the synoptic environment of tropical cyclones.

Many examples of the utility of model-assimilated data sets in research can be cited. Most of the published work on diabatic forcing of the atmosphere has been accomplished by groups at the universities of Wisconsin, Utah, Reading, and Helsinki using model-assimilated data sets. Most of the diagnostic work on blocking, multiple flow regimes, and low-frequency variability undertaken at NOAA/NMC's Climate Analysis Center (CAC); the Massachusetts Institute of Technology; the universities of Yale, Washington, Stockholm, and Bologna; and the ECMWF have used model-assimilated data sets. Much of the diagnostic work on the structure of the 30-to 60-day oscillations at the universities of Honolulu, Wisconsin at Madison, California at Los Angeles, and Tsukuba and at the NCAR and CAC has been based on model-assimilated data sets, frequently in conjunction with satellite measurements of outgoing long-wave radiation (OLR). The group at the University of Reading, in a recent climatological atlas based on ECMWF model-assimilated data sets, cited 32 papers "published from or inspired by" their climate diagnostics project over the last decade. Groups at ECMWF and the universities of Cologne, Kiel, Wisconsin, and Helsinki have extensively documented the atmospheric energy cycle using model-assimilated data sets. Finally, model-assimilated data sets have proved invaluable for oceanographers concerned with forcing of the oceanic circulation. The data sets provide a coherent synthesis and interpretation of all available information on the surface fluxes of momentum, heat, and moisture that drive ocean circulation models and ocean wave models.

4

Needs for the Future for Model-Assimilated Data Sets

RESEARCH PROBLEMS

Current research problems that require the use of model-assimilated data sets in the future are summarize in this chapter. The concise summary includes a discussion of the temporal and spatial scales encountered and the implied resolution of the data fields that need to be produce and archived.

Synotic Systems

A thorough understanding of the structure and dynamics of synoptic weather prediction (NWP). While the study of baroclinic instability and wave cyclones has been a central theme of dynamic meteorology for over 40 years, the generation, maintenance, and decay of these synoptic weather system, as well are still far from completely understood, and new field programs for their study are being organized. The horizontal an vertical structure of the synoptic system, as well as the relative importance of dynamic and thermodynamic mechanisms in their life cycle, are topics of current active research. The preferential location of storm tracks and their variability connects cyclonic scale to spatially and temporally longer scales, which are discussed in the next two sections.

The characteristic length scale of these phenomena is $L = 10^3$ km, although frontal structures with cross-front $L = 10^2$ km play an important role.

during periods of maximum amplification. Their characteristic time scale is T = 3 to 10 days, with the first 2 to 3 days constituting the growth, or amplification, stage.

Planetary Waves

Planetary waves have a crucial role for medium-and extended-range forecasting (MRF, ERF). In midlatitudes, Rossby and inertia-gravity (Poincaré) waves play a major role, while the tropics give rise to Kelvin and mixed Rossby-gravity (Yanai) waves. Aside from the troposphere, these waves also appear, with various degrees of importance, in the middle atmosphere and oceans. Their horizontal and vertical propagation and their interaction with the mean flow are major topics of interest.

Spatial scales of planetary waves in the atmosphere are $L = 10^3$ to 10^4 km, and temporal scales are T = 3 to 30 days, except for Poincaré waves, which are characterized by T = 1 to 12 hr. In the ocean the corresponding spatial scales are shorter and the temporal scales are longer.

Low-Frequency Variability

The predictive counterpart of the low-frequency variability (LFV) is long-range forecasting (LRF). Important research issues concerning LFV include barotropization, that is, the increasingly barotropic character of atmospheric motion as the characteristic period increases, the existence and explanation of multiple flow regimes for the same external conditions, and the emergence and life cycle of localized coherent eddy structures (Ghil and Childress, 1987). Various intraseasonal oscillations, with dominant periods of 16, 25, and 40 to 50 days, have been observed in the tropical atmosphere as well as in the northern and southern hemisphere extratropics. These oscillations need to be explained and related to tropical-extratropical interactions and to energy and momentum transfer by wave dispersion. The length scale of phenomena in this category is $L = 10^3$ to 10^4 km, and the corresponding time scale is T = 30 to 300 days.

Atmosphere-Ocean Interactions

A brief list of important research problems in this category includes the tropical 40-to 50-day propagating wave disturbances and their role in El Niño-Southern Oscillation (ENSO) events, the genesis and life cycle of sea surface temperature (SST) anomalies, explosive marine cyclogenesis, and the dependence of planktonic biota on upwelling and hence wind stress. These phenomena span the time scale of the previous three categories, with $L = 10^2$ to 10^4 km and T = 3 to 300 days.

Atmosphere-Land Interactions

Some relevant problems in this area are persistent droughts and severe winters. Changes in surface heat and momentum fluxes with these events are associated with increased snow and ice cover—in the latter case leading to enhanced albedo and in the former to increased evapotranspiration from land vegetation. Typical spatial and temporal scales of interest here are L = 100 to 1000 km and T = 10 to 300 days.

Dynamics-Chemistry-Radiation Interactions

Rather than occurring at a two-dimensional interface, as for the two preceding types of interaction, dynamics-chemistry-radiation interactions occur throughout the depth of the lower and middle atmospheres. The vertical scale can vary from hundreds of meters to tens of kilometers, while the range of horizontal and temporal scales covers the entire spectrum discussed so far.

Hydrology

The development of comprehensive and consistent model-assimilated data sets for the global hydrological cycle requires a coupled treatment of the various subsystems. This requires a capability for assimilating the current highly nonhomogeneous hydrological information as well as the remotely sensed information planned for the future. Such a system seems to be best conceptualized in terms of a real-time operational component and a delayed research-quality climatic database component.

The future space-based system for precipitation estimation will rely on a variety of indirect measurements (e.g., visible and infrared passive microwave radiances), along with direct estimates from precipitation radar (Tropical Rainfall Measuring Mission Science Steering Group, 1988). Integration of this information with information from in situ measurements and model output to obtain optimum global precipitation fields win be a difficult task.

The advantages of a comprehensive analysis system for the hydrological cycle are well illustrated by considering the common exchange quantity of evaporation minus precipitation (E-P), that is, the net transfer of water substance from the earth's surface to the atmosphere. Global atmospheric forecast models, through the data assimilation process, should ultimately provide accurate estimates of temporally and a really averaged vertically integrated vapor flux divergence E-P over most areas of the world. Accurate evaluation of the flux divergence is most difficult over areas of sparse data and regions of significant relief, where smoothed model terrain may bias estimates. These estimates can then be compared with the difference between parameterized evapotranspiration and "measured" precipitation, and any inconsistencies can be resolved. The reconciled data set can then be viewed in the context of the evolution of the ocean thermal fields to resolve apparent inconsistencies in these two subsystem analyses. Finally, the atmospheric estimate of E-P over basins where streamflow is also measured can be used to derive the month-to-month changes in surface and subsurface moisture storage, quantities that at present are only roughly known and are of extreme importance in the modeling of climate change (Rasmusson, 1968). Variations in E-P over ocean areas are important in generating systematic and white noise components of forcing on the ocean circulation. Thus, we need to know the synoptic distribution of E-P over the ocean as well as overland.

DATASETS NEEDS BY TIME SCALES

The partial differential equations governing atmospheric and oceanic motions are nonlinear. Therefore, different temporal and spatial scales interact with each other, and no particular frequency or wavenumber band can be completely understood without consideration of the adjacent bands or even without the help of much shorter or much longer scales. Still, for convenience, the discussion in this section is subdivided into frequency bands.

0 to 3 Days

The characteristics of the meteorological data systems of the 1990s will require the use of analysis procedures that are more complex than those currently in use. For example, the continuous nature of some of the data sources will require a higher-resolution analysis system. Also, the varied error characteristics of the data will need to be objectively accounted for. Regardless of spatial or temporal scales, the use of models in data assimilation is motivated by the need to impose dynamic consistency on the data sets, to interpolate the data in some optimal way to a grid system, and to provide reasonable estimates of meteorological structures where there are large spatial or temporal voids in the data. An additional motivation for using models for mesoscale data assimilation is that the resolution of conventional operational data has often been insufficient to define the detailed characteristics of mesoscale meteorological processes. Use of fine-grid models, with appropriate mesoscale physics, is necessary in order to generate a data set for subsequent analysis. In this case, the model is used to *simulate* mesoscale data as well as to integrate observed data. While this particular contrast with the objectives of large-scale data assimilation is not likely to be completely eliminated in the future, new data acquisition systems such as the WSR-88D Doppler radar and the wind profilers will allow much better routine identification of mesoscale structures than is now possible. However, since many variables will still be poorly defined, research-oriented data assimilation systems that are tailored to specific scales, geographic regions, or physical processes will continue to be required.

3 to 10 Days

Although objective analysis and numerical prediction originated from the demand for synoptic weather forecasts of 1 and 2 days, the scope of forecasts has gradually extended to the medium-range time scale of 3 to 10 days. Simultaneously, analysis techniques have recognized the importance of using numerically predicted fields as a background field for analysis, and of requiring dynamical consistency among variables, resulting in the continuing development of data assimilation methodology. The models used as assimilators have mostly been global general circulation models (GCMs), as opposed to limited-domain models for short-range forecasts. Assimilation techniques for medium-range forecasts have now reached a fairly mature stage of development, but needs for further development in the future can readily be identified.

These needs may be grouped into two categories. One is the correction of current known deficiencies, and the other is general improvement of data assimilation systems for the purpose of achieving further advancement in medium-range forecasting, particularly beyond about 5 to 6 days (Hollingsworth, 1987).

The deficiencies vary depending on the assimilation system. One of the outstanding deficiencies often mentioned with respect to the medium-range forecasts of operational centers is the "spin-up" problem. There are two aspects to this problem: one is of a pathological nature resulting from lack of adequate model initialization, and the other is due to discrepancies, often noted as "systematic errors," between a model solution and observations.

Development of the nonlinear normal-mode initialization technique has substantially improved initialization; as a result, pathological problems in the extratropics have been dramatically reduced. However, tropical initialization still is imperfect in terms of including diabatic heating correctly and therefore is imperfect in handling condensation and evaporation rates. For example, it takes 2 to 3 days after the beginning of a weather forecast for the rates of precipitation and evaporation to reach equilibrium. Some scientists believe that this drawback can be eliminated only by both better parameterization and "genuine" four-dimensional analysis that employs the adjoint method or the Kalman filter (see Chapter 2). On the other hand, other scientists consider that appropriate subgrid-scale parameterization by itself can alleviate this problem. Others consider that continuous data assimilation, including the adjoint method and the Kalman filter, can solve the spin-up deficiency.

Rectification of systematic errors in the assimilating model will also reduce errors in the model-assimilated data sets, particularly for data-void areas. However, reduction of a GCM's systematic errors may require a long, careful effort. Systematic errors, such as a global cooling tendency, a meridional shift of westerly jets, and an excessive intensity of easterlies at the tropical tropopause level, have long been noted, but clear-cut remedies have not yet emerged. In fact, such biases in models may be most efficiently corrected by interactive use of data assimilation and model predictions, as discussed in Chapter 1.

The needs in the second category, general improvement of data assimilation systems, include better assimilation algorithms, more comprehensive observational networks, more reliable observational platforms, and improved quality control of observational data. While some observational data problems are instrumental, the discussion here is concerned with problems in which improvement can be achieved using observed data interactively in an assimilation system. Such tasks include (1) utilization of outgoing long-wave radiation (OLR) data; (2) better use of improved satellite temperature retrievals, cloud winds, and precipitable water estimates; (3) use of future scatterometer data; and (4) improvement of equatorial surface wind analyses.

Use of OLR data has been suggested for producing better initial tropical wind divergence fields (Julian, 1984; Krishnamurti and Low-Nam, 1986; Kasahara et al., 1988). The next step is to investigate whether such usage improves model rainfall forecasts and whether enhanced or rectified cumulus convection can be maintained by data assimilation without need for a spin-up by the model.

Satellite data have been used for temperature retrievals since 1968, for cloud winds since 1970, and for moisture since 1983. The benefits of these data for improving forecasts have been well demonstrated in several numerical experiments (Uppala et al., 1985; Kalnay et al., 1985; Illari, 1989). Some of these retrieved data have been used routinely in operational data assimilation, but questionable data are still occasionally reported. Raw radiance data, rather than retrieved temperatures, have been proposed for direct assimilation since radiances are what satellites actually measure. Cloud wind data continue to be plagued by a nagging problem of proper height determination. Improvement in locating cloud levels is highly desired.

Scatterometer data are expected to be available in the near future, and the inclusion of such data in data assimilation systems may contribute significantly to better representation of surface wind stress. Based on an assimilation experiment using a limited data set from the Seasat-A scatterometer, Atlas et al. (1987) recommend that the directional accuracy of surface winds from future scatterometers be substantially improved.

The quality of equatorial wind analyses has been studied intensively by comparing analyses from various operational centers with the 1997 surface wind data of moored buoys (Reynolds et al., 1989). The results are disappointing, somewhat surprisingly, since the operational wind analyses were thought to be acceptable. The quality of these data is important for modeling long-term events like El Niño as well as for day-to-day tropical variations.

10 to 100 Days

It appears appropriate to subdivide this time range into two ranges (i.e., 10 to 30 days and 30 to 100 days) from the standpoint of atmospheric as well as oceanic forecasts. The forecasts in the first range, 10 to 30 days, are often referred to as 30-day or 1-month forecasts; those in the second range, 30 to 100 days, are referred to as seasonal forecasts. The former has been an object of intensive studies (e.g., dynamic extended-range forecasts [DERF]) at a number of operational centers around the world (National Research Council, 1991). For forecasts in this category, reliable data assimilation of large-scale atmospheric circulation features is required to produce appropriate initial conditions. In addition, accurate observed sea surface temperatures must be specified at the initial times.

A fundamental consideration is that 30-day forecasts are not deterministic because the limit of predictability has been exceeded and that only probabilistic forecasts are possible. This implies that some forecasts among members of ensemble forecasts are good, while other members are not good, even in the case of a perfect forecasting model. A crucial problem is how to generate ensemble forecasts based on multiple initial conditions. The model-assimilated data sets are quite relevant to this issue. For example, Hoffman and Kalnay (1983) proposed an approach referred to as lagged average forecasting. Another well-known approach is to generate various initial conditions by adding random numbers to a basic initial condition; this is called the Monte Carlo method (Leith, 1974). Current ideas under investigation include finding the fastest-growing mode of the symmetric eigenvalue problem.

The forecasts in the second category (30 to 100 days) require coupled air-sea models for prediction, and the initial conditions should be produced by data assimilation systems based on the coupled model. The current state of ocean data assimilation is discussed in Chapter 3. There have been several studies on global ocean data assimilation, but no attempt at coupled data assimilation has yet been conducted. Two essential elements in coupled assimilation are that the ocean heat content should be properly included and the surface winds accurately represented. These are very important because El Niño forecasts critically depend on the distributions of these variables in the initial condition.

The data required for this purpose are, in addition to the conventional atmospheric data, the surface winds, the surface ocean temperature, and the subsurface temperature. Furthermore, it is envisaged that altimeter data and scatterometer data from satellites could considerably enhance more conventional observational systems. The Tropical Ocean and Global Atmosphere (TOGA) program is testing the utility of drifting buoys, which provide surface ocean current data. TOGA also is experimenting with moored buoys, which provide ocean temperature, salinity, and subsurface current observations. These data are potentially valuable and important.

100 to 1000 Days

On seasonal-to-interannual scales, the value of atmospheric model-assimilated data sets has already been demonstrated (see Wallace, 1987, for a review and a rather complete set of references). Wallace and his collaborators have used 15 years of 12-hourly National Meteorological Center (NMC) model-assimilated analyses to study low-frequency variability in the northern hemisphere atmosphere poleward of 20°N. Resulting discoveries include Pacific-North America (PNA), North Atlantic, and North Pacific oscillations (Wallace and Gutzler, 1981), as well as the correlation between the PNA oscillation and the Southern Oscillation index (Horel and Wallace, 1981). Numerous other studies of variability on slow time scales have been performed using the NMC model-assimilated data set—studies that could not have been done using the rawinsonde station data directly.

While current operational analyses of atmospheric data have flaws, omissions, and (largely undocumented) discontinuities in data assimilation procedures, it is recognized that data assimilation techniques have improved markedly, especially in the tropics, and that these model-assimilated data sets have proven extraordinarily valuable in examining the low-frequency variability of the global atmosphere. It has been suggested that model-assimilated data sets be periodically reanalyzed by NMC and other operational or research centers to include new or omitted observations as improved data assimilation techniques and more accurate assimilation models are developed. The advantage of such reanalysis would be to provide the highest-quality global atmospheric data sets possible with which to perform additional studies of low-frequency variability.

In the ocean the primary need for model-assimilated data sets is to initialize the ocean for coupled atmosphere-ocean model predictions and for integration of the results of oceanographic field programs. In order to make predictions on time scales of months to years, the evolution of sea surface temperature (SST) must also be included, and the only way to do this is to make forecasts using coupled atmosphere-ocean models. The work of Cane et al. (1986), using a simplified SST anomaly coupled model, has indicated that the onset of the ENSO phenomenon in the Pacific is potentially predictable as much as a year in advance and that the major source of forecast error is specification of the initial state of the ocean. The model currently obtains an initial ocean state for the forecast by forcing the ocean to equilibrium with the observed winds and allowing the model to freely evolve afterwards. No ocean data are used to initialize the model; therefore, errors in the observed winds can lead to errors in the initial specification of the ocean state. A more comprehensive routine effort at ENSO prediction using coupled GCMs would require the assimilation of ocean data and would produce ocean model-assimilated data sets as a by-product.

Note that once the ocean and atmosphere are coupled completely, correct simulation of the annual cycle is not guaranteed. Accurate simulation of the annual cycle thus becomes a crucial test of coupled models.

With the exception of the work in the Atlantic and Pacific cited previously, production of oceanographic model-assimilated data sets is almost nonexistent. Oceanographic field programs, such as TOGA, the Seasonal Equatorial Atlantic Experiment/Français Ocean et Climat dans l'Atlantique Equatorial (SEQUAL/FOCAL), and Tropic Heat, that extend over many years have been performed in the tropical oceans; similar programs, such as the World Ocean Circulation Experiment (WOCE) and the Joint Global Ocean Flux Study (JGOFS), are being started for the extratropics. Because the observed data in these programs were of different forms, what has emerged is a collection of different data streams (e.g., drifter, current meter, expendable bathythermograph (XBT), satellite SST, satellite altimeter), all describing aspects of the ocean but not dynamically connected and therefore not mutually reinforcing. Assimilating all these data into an ocean GCM would (1) provide a general description of the interannual variability of the ocean; (2) provide a data set for the initialization of coupled ocean-atmosphere predictions; and (3) provide consistent dynamical quantities, such as vorticity and vertical velocity, that cannot be measured directly but that can be computed in the process of assimilating the data.

1000+ Days

At time scales beyond a few years, small effects acting over long periods of time can have major climatic impacts. Simulating and predicting large-scale climate change on these time scales involves many physical, biological, and chemical aspects of the atmosphere, ocean, cryosphere, and land system. In order to simulate and predict climate changes on long time scales, a model with all these elements and their interactions needs to be constructed. While much effort has gone into part of this problem, a full-blown model has not yet been developed.

The instrumental observational record is limited to little more than about a hundred years at the surface and less than 40 years at upper levels of the atmosphere. Within the ocean the record is uneven and sparse; in some regions of the ocean, deep measurements have never been taken.

The problem on long time scales is therefore twofold. There is a data problem of combining individual records into global fields of data over long time intervals and a modeling problem of devising accurate and comprehensive enough models to be able to simulate and predict the global climate over decades and longer.

In the atmosphere an upper-air network has been used as the starting point for routine global weather forecasts since the mid-1950s. The analyses of upper-air data constitute a long record of global atmospheric model-assimilated data sets but one that is inadequate for climate purposes because it is so spatially and temporally inconsistent. Several changes in the data collection system and the analysis techniques mask long-term climate variability and trends that exist in the data. As pointed out previously, the existing record of weather analyses is good enough to show shorter-term variability but is totally inadequate for changes on scales of a few years to decades.

A remedy, simple in concept but difficult and expensive to execute, is to analyze the existing 40 or so years of original data in a uniform manner using the best available assimilation models. This would produce the best possible long-term model-assimilated data set of the global atmosphere, one that would then be useful for analysis of long-term climate fluctuations. Since, in the future, this type of data set would be the most useful long-term global atmospheric data set in existence, systematic and consistent additions should continue the archive into the future in order that reanalysis of the data set can be accomplished later with advanced assimilation models. In any reanalysis, data that were not available at the time of the previous analysis could be profitably included.

The ocean situation is quite different in that data do not exist to define the evolution of the ocean system over periods of years and beyond. It is important, however, to have the *mean* state defined, since many of the changes that the ocean will undergo can be calculated by moving heat, momentum, and constituents around by the mean circulation of the ocean. To the extent that the ocean is in steady equilibrium (i.e., the eddies are statistically stationary and the annual cycle is regular), the best way to define this state would be to assimilate the existing data into a seasonally varying model of the global ocean.

Making predictions on time scales of a few years to beyond a hundred years requires careful initialization of the ocean state. There is evidence that the ocean SST affects climate on decadal time scales (e.g., Palmer, 1986) and that part of this SST variability is related to long-term changes in the deeper parts of the northern oceans (e.g., Lazier, 1988; Levitus, 1989).

Unless the initial state can be specified, the future evolution of the ocean cannot be predicted. The time range over which specification of the initial ocean state affects details of the final climatic state is undetermined at present. On time scales of thousands of years, the coupled climate state presumably loses all memory of the initial ocean state. In order to determine the influence of the initial ocean state, climate runs of coupled global models starting from slightly different initial conditions in the ocean must be done. Once the range of deterministic predictability has been estimated, the requirements for initializing predictions may be better defined.

REQUIREMENTS FOR DATA AND COVERAGE

General Principles

To study a phenomenon of given length scale L and temporal scale T, the required resolution is roughly L/10 and T/10. The horizontal extent of data coverage for all the phenomena discussed in the preceding section should be global. For some of the phenomena, regional, hemispheric, or tropical-belt coverage might be satisfactory for some applications, but global coverage is needed for others. Regional data sets with higher resolution can be very useful complements to global sets that satisfy minimal resolution requirements. Such high-resolution data sets should cover at least two separate regions for comparison purposes. Horizontal extent of regional data sets should generally be at least 10L.

The vertical coverage should be from the land and ocean surface to at least the mesopause. For some of the phenomena discussed, the coverage should extend downward well below the ocean surface and somewhat below the land surface. The temporal coverage should extend over 10T to 100T in order to achieve statistical significance and permit analysis of slow changes in the phenomena of interest.

General Specifications

Horizontal and vertical resolution, as well as temporal resolution, will vary among global-scale, regional-scale, mesoscale, and local-scale data sets. In general, it appears at this time that the following general specifications represent an achievable goal.

Most data sets assimilated and archived should have global coverage, with a horizontal resolution of 100 x 100 km or better. The vertical resolution should be 30 levels judiciously distributed, that is, with greater resolution in the subsurface, the planetary boundary layer, and the lower stratosphere.

Selectively increased horizontal and vertical resolution in certain regions should be employed in the coming decade as mesoscale data become routinely available from the modernization program of the National Weather Service (NWS). Increased resolution is also necessary for the purposes of designing or verifying the results of field programs. A recommended resolution for such selected regions is 10 x 10 x 50 km levels.

A data set of barely satisfactory length for the study of low-frequency variability and atmosphere-ocean interactions, atmosphere-land interactions, and dynamics-chemistry-radiation interactions is 30 to 40 years. Such model-assimilated data sets need to be as temporally consistent as possible.

5

Quality Control and Validation of Observations, Analyses, and Models

As discussed in Chapter 1, the application of data assimilation includes a structured open-ended learning process that improves our understanding of both the factors governing the time-dependent evolution of a geophysical fluid system and the characteristics of the data. This learning process occurs through the many ways in which different data streams are systematically compared with each other and with predictions of short-range fore-casts from the assimilating model (Hollingsworth, 1989). Inconsistencies between the observations and predictions are easily documented and demand explanation, providing the basis for quality control and validation of observations, analyses, and the models themselves.

DATA MONITORING

Significant improvements over the last decade in the use of data assimilation for numerical weather prediction (NWP) have resulted in the development of new techniques for global validation and quality assurance of both in situ and remotely sensed data (Hollingsworth et al., 1986).

Modern data assimilation systems use the relevant prior data and a state-of-the-art computer model of the atmosphere to provide an accurate quantitative estimate of the current state of the atmosphere. For each and every observation of a variable that is made, a background value derived from the model forecast is also available for comparison. In data-sparse areas, systematic differences between the forecast values and the observations are frequently indicative of systematic errors in the observations. In data-rich areas the observations and the forecast values have similar accuracies on the scales resolved by the forecast model, and the accuracy of the forecast is usually good enough to be very useful in detecting uncharacteristically large errors in the observations. However, there are instances in which the model has rejected valid observations representing a significant change that result in a subsequent incorrect forecast. Methods have been developed to recognize these critical events, but it cannot be said that state-of-the-art models have entirely resolved this problem.

The relatively high accuracy of short-range forecasts is used by NWP centers to monitor systematically many types of observations. Such monitoring has successfully identified long-standing errors in radiosondes at remote stations, in reports from ships plying remote routes, in aircraft reports over the oceans, in cloud-track winds from geostationary satellites, and in temperature soundings from polar-orbiting satellites. As the NWP systems have improved, they have exposed significant defects in current satellite ground-processing schemes and limitations of current instrumental technology. Serious problems of bias and noise have been uncovered in retrievals of wind and temperature (Andersson et al., 1991; Kelly et al., 1991).

In many cases the analyzed fields provided by the assimilation system can be verified directly against in situ measurements. Thus, the model provides the guidance to identify problems, and carefully chosen in situ measurements provide conclusive proof.

EPISODIC MODEL REJECTION OF DATA

The relationship between data that must be quality checked and the background field that the data are used to correct is somewhat paradoxical. That is, the background field generated by the model must be sufficiently accurate to allow detection of erroneous data. However, these model-generated background fields contain significant errors, especially in data-sparse regions—if these errors did not develop in the forecast background field, there would be no need for data to correct them. Thus, a decision whether to correct a background field based on a current datum or to reject the datum because it differs by some increment from the prior estimate is difficult.

This problem is complicated by the fact that even correct data should sometimes be rejected if they reflect processes that cannot be resolved on the scale of the grid system used in the analysis. For example, an observation may correctly reflect the existence of a local thermal or velocity field anomaly associated with a thunderstorm outflow boundary; however, it would be undesirable if this datum were allowed to modify a background field on a grid with, say, grid boxes of 100 km on a side, because the datum is not *representative* of the scales resolved by the model grid and its influence would be distributed over an erroneously large area. Thus, sometimes even good data need to be rejected by the assimilation system.

The uncertainty that stems from an inability of automated quality control of data to deal with all situations encountered requires follow-up human evaluation and intervention. Human intervention at times includes reintroduction of information and previously rejected data into the analysis by more subjective procedures. Ideally, an experienced analyst would recognize a poorly observed atmospheric process that is not represented well by the background field and subjectively modify the analysis to ensure consistency with a conceptual understanding of the process or phenomenon on the scale of the model. This is an important area of research, particularly for mesoscale data assimilation where sometimes few data points are available to define important mesoscale structures.

VALIDATION OF REMOTELY SENSED EARTH OBSERVING SYSTEM DATA

Data from any new or existing satellite observing system will not be really useful unless the errors in the data are comparable with or smaller than the errors in prior information, as represented by the model-produced background values. This is just as important in the use of satellite data for climate studies as for NWP. Sustained and intensive monitoring, quality assurance, and global validation of the algorithms and data from a new system such as the Earth Observing System (EOS) are essential to ensure that the observations from the system are of sufficient quality to be useful for both climate and NWP analyses.

The importance of a global approach to validation of remotely sensed satellite data has been demonstrated by operational experience with cloud track winds and temperature soundings. For example, wind data from the satellite Seasat-A scatterometer had many anomalies and biases. Seven years after launch, data assimilation studies of the scatterometer data documented the already-known biases and identified others. Both sets of biases were documented conclusively using the Seasat-minus-forecast and Seasat-minus-ship comparisons generated by a few days of data assimilation (Anderson et al., 1991). The biases would have been detected much sooner if the Seasat data had been critically evaluated in a data assimilation system.

The distinction between research satellite missions and operational satellite missions is losing its significance. To produce research-quality data from a new satellite system, the observed data should be subjected to a critical evaluation by an assimilation system in order to identify error characteristics of the instruments and the algorithms. If an operational NWP system has predictive capability for the remotely sensed quantity, the real-time assimilation system can provide a basis for validating and assimilating the new data. The data assimilation system then provides a powerful and systematic means for comparing the new remote measurements with all earlier and current in situ and remotely sensed measurements. Thus, the real-time operational assimilation system can provide quality assurance and validation of the new satellite observations.

Experience with many satellite systems shows that real-time assimilation systems at NWP centers provide rapid identification and diagnosis of problems that would otherwise pass undetected for long periods, thereby corrupting irreplaceable data. Real-time indications of sudden problems can be provided within a few hours, as erroneous data are rejected by the assimilation system's quality control programs. In these checks, "toss-out" criteria are applied at two stages in the comparison of background field and observed values.

In addition, the real-time assimilation system provides a comprehensive, quantitative "quick-look" synthesis on a regular grid of all current and past observations. Oceanographers, for example, will require most EOS atmospheric data in assimilated form on a regular grid as input to ocean wave and ocean circulation models. Early availability of real-time operational analyses will stimulate research on new satellite observations and demand for more rapid production of delayed-mode analyses.

The use of EOS data in operational data assimilation will therefore be the first iteration of the research use of the data, and so the operational centers can contribute a great deal to the success of the overall research goals provided they are tasked and funded to produce research-quality assimilation data sets during the daily cycle.

VALIDATION OF MODEL-ASSIMILATED ANALYSES

Many methods have been used to validate the description of the atmosphere provided by model-assimilated data sets. The simplest method is to calculate the mean and root mean square differences between the analyses and the observations used in the analyses. Any analysis scheme ought to be able to fit the data used to within reasonable bounds (Hollingsworth et al., 1985).

A somewhat more stringent method, which can be applied in data-dense regions, is to withhold some of the observational data from the analysis procedure and to validate the analyses against the withheld data.

A method much favored by NWP researchers is to validate the analysis through observational verification of very short range forecasts made from the analyses. The random component of observation error is independent of the forecast, so an estimate of the forecast error, which is an upper bound on the analysis error, can be obtained. Since the short-range forecast usually provides the background field for the next analysis, this approach is equivalent to observational verification of the background field (Hollingsworth and Lönnberg, 1986).

All of the approaches thus far involve calculations at single points. A more revealing approach is to examine two-point correlations of the departure between analyses and observations. This is a very efficient method to determine if the analyses have in fact extracted all the available information from the data. If the calculations are multivariate (involving, say, wind or wind shear at one point and geopotential or thickness at another), one can readily determine if the analyses reflect the balance of the observations (Hollingsworth and Lönnberg, 1989).

The vertically integrated latent and sensible heating of the atmosphere estimated from model-assimilated data sets can be validated in a number of ways. Measurements of rainfall can be used to validate the vertically integrated latent heat release. The sum of the net radiation at the top of the atmosphere and the atmospheric heating gives the net flux into the underlying land or ocean surface, which can be measured or estimated in various ways. In the tropics, outgoing long-wave radiation (OLR) data (which is still not used operationally) is useful for validating inferred vertical velocity and diabatic heating fields in the tropics (Arpe, 1991). Ultimately, satellite-borne active and passive microwave, measurements will provide more accurate measurements of moist diabatic processes.

Diagnostic studies have deduced many highly derived quantities from model-assimilated data sets. Intercomparisons of these results for Global Weather Experiment (GWE) data showed large discrepancies, indicating that some or all of the analyses were unreliable. Over the last 7 years, successive reanalyses of GWE data have produced a marked convergence in the results, although there is still some way to go. The atmospheric energy budget is a prime concern of several component programs of the World Meteorological Organization's (WMO) World Climate Research Program (WCRP), especially of the Global Energy and Water Cycle Experiment (GEWEX).

A validation method of increasing importance is the use of atmospheric model-assimilated data sets to drive ocean circulation or ocean wave models. Both models are controlled by atmospheric forcing and are extremely sensitive to it (Harrison et al., 1989; Janssen et al., 1989). The responses of ocean models to differing atmospheric forcing are large enough to be easily verified against oceanographic observations. Such ocean models provide useful tests of atmospheric model-assimilated data sets.

The most comprehensive validation of model-assimilated data sets is the skill of daily forecasts. Forecast skill has improved markedly over the last decade, largely as a result of increased analysis accuracy.

VALIDATION OF FORECAST MODELS

An important application of model-assimilated data sets is validation of the physical parameterizations used in general circulation models. Model-assimilated data sets reveal the close ties between model parameterization schemes; the initial tendency for mean errors; and the fully developed mean errors after many days of integration by tracking energy, momentum, and other balance requirements for the atmosphere.

The "balance requirement" approach has been used for decades, especially in the GWE data, to estimate atmospheric diabatic forcing by means of the fact that the time mean tendency of the atmosphere is zero on time scales of a month, apart from a small and easily calculated seasonal trend. The observed balance of the atmosphere is used to infer the mean diabatic forcing of the atmosphere through tendency calculations that apply the adiabatic equations of motion to the analyzed data. From the principle of balance, the average diabatic forcing is the negative of the average adiabatic tendency. New applications of this principle have recently been found in the validation of the parameterizations used in models.

Calculations of the monthly average of a model's initial adiabatic tendency and diabatic tendency then provide three-dimensional fields of the "true" diabatic forcing and of the errors in the model's diabatic forcing. Recent applications of this methodology to model-assimilated data sets and to 1-day and 10-day forecasts have shown that there are close similarities between the mean errors (sometimes referred to in jargon as climate drift) evident in 1-hour, 1-day, and 10-day forecasts.

Comparisons of such results for data sets assimilated with models differing only in their physical parameterizations have been of great value in documenting differences in performance of the parameterizations and in linking the differences in parameterization to differences in mean error evolution.

This new approach to validation of parameterization schemes is important for their successful development. Hitherto, parameterization schemes have been based on conceptual understanding of physical processes supported by field experiments and detailed modeling. The field data for process studies are frequently combined into a time-evolving single-column composite description representing mean values of an areal observing network. As a result, parameterization schemes generally have been developed in a one-column atmospheric context and validated on one-dimensional field data.

The balance requirement approach to validation of parameterizations uses a "top-down," rather than a "bottom-up," approach to model validation. Using copious quantities of operationally available data that describe synoptic-scale phenomena, the method asks, what do the synoptic-scale and large-scale motions need from the parameterization schemes in order to maintain the observed balance? The approach has proved valuable in diagnosing problems in model formulations of gravity wave drag near mountains; in radiative, convective, and planetary boundary layer schemes used in models; and even in analysis methods.

Of course, this new approach only indicates that a problem exists in certain terms in the equations; it does not of itself solve the problem. Improvement of a parameterization scheme within a forecast model is a delicate task, since parameterization schemes can produce complex dynamical effects on the evolution of the flow in a forecast model. However, examination of the very short range forecast errors is an effective first step in diagnosing errors in the physical parameterizations of a model. Deficiencies in the interactions between the dynamics and the parameterization schemes can be difficult to identify, so a thorough understanding of how each physical process operates is required in order to suggest possible improvements. This new approach to model validation provides useful insight into the sources of errors in the physical parameterizations, and it provides a valuable additional method for systematic assessment of the performance of model physics. Its effectiveness depends entirely on the availability of the model-assimilated data sets.

6

Status of Data Archives, Access, and Future Directions

In this chapter the status of available archived model-assimilated data sets (MADS) and supporting observed data is discussed, including the flow of data to archives and the accessibility of data by users. The need for a nationally focused and integrated archive system for assimilated data sets with provision for ready access is discussed. A future view of access to data sets, including software needs, also is presented.

AVAILABLE ANALYSES AND OBSERVATIONS

Only a brief outline of available observational data and model analyses is attempted in this report. An in-depth effort is needed to locate, evaluate, and compile available geophysical data resources and model analyses prior to development of a state-of-the-art model-assimilated data set extending back to about 1950, as recommended by the panel.

Atmospheric Analyses

Some surface analyses exist for the years since 1900. Archives of some upper-air analyses are available for the years since 1946. Daily model assimilation products exist for the years since about 1960. A daily set that includes ocean surface flux terms and model radiative terms is available for the years since 1985.

Ocean Analyses

Various analyses of surface fields (sea surface temperature [SST], stress, etc.) exist for 20 years to a century. Model-assimilated data have been available only for a few years and are not yet global. However, global ocean MADS, starting about 1985, are expected to become available in 1991.

Hydrological Analyses

River basin simulations have been run for more than 10 years for the United States. To the panel's knowledge, the fields have not been saved.

Paleoclimate Reconstructions

"Observed" fields such as topography, ice extent, and SST for the peak of the ice age (18,000 years ago) have been prepared for the world. Several climate models have been run using these fields as boundary conditions.

Surface Observations (Meteorology, Hydrology, Oceanography)

Archives of surface marine observations exist for the years 1854 to the present, with gaps during the two world war periods. A major project, the Comprehensive Ocean-Atmosphere Data Set (COADS), involving cooperation among the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration/Environmental Research Laboratories (NOAA/ERL), the National Climatic Data Center (NCDC), and the University of Colorado's Cooperative Institute for Research in Environmental Sciences (CIRES), is under way to improve this data set. The main world archives of surface land data sets begin with the year 1967. A considerable amount of earlier data on tape also exists in diverse locations. In many areas, surface observations were routinely recorded at least by 1880, but the data would be difficult to compile and digitize.

The U.S. Geological Survey maintains a very good archive of daily and monthly river flow data for the United States. Monthly river flow data exist for several hundred sites in other parts of the world, but global coverage is far from complete.

Upper-Air Data (Meteorology)

Archives of global sets of daily upper-air rawinsonde observations start with the year 1957, with a gap in much southern hemispheric data from 1963 to 1966. Data from Australian and U.S. rawinsonde stations are archived for the years from the late 1940s to the present. Archiving of satellite sounder data started in 1969, but there is a gap for 1971–1972.

These archives provide starting points for a state-of-the-art assimilation analysis effort. The archives exist because of patient, low-cost data-gathering efforts at a few centers. For purposes of reanalysis, archives should include data from delayed ships and isolated locations that were not included in real-time data sets. These efforts need enhancement. All available world data resources should be brought together in a nationally focused effort to develop an assimilation data archive system.

RESOURCES NEEDED FOR ARCHIVING MODEL-ASSIMILATED DATA SETS

New model assimilation products will continually become available. Additional efforts are needed at the NCAR, the NCDC, and the National Oceanographic Data Center (NODC) for efficient archiving of MADS. In addition, major MADS producers such as the National Meteorological Center (NMC), the European Centre for Medium Range Weather Forecasts (ECMWF), and others will need to apprise the archival centers of significant model attributes, accessing and imaging guidelines, and changes associated with the data.[1]

Resources Needed to Prepare Older Observed Data

Major efforts will be necessary to prepare older observed data for input to new analyses. An enhanced effort is needed in several discipline areas: marine ship data, upper-air meteorological data, and many types of land surface data (synoptic, solar, river discharge, soil moisture, etc.). An enhanced effort is needed for data sets that span more than a decade. International cooperation is essential; some of the needed coordination can be handled by the World Climate Research Program (WCRP) and bilateral programs.

ARCHIVE METHODS AND INSTITUTIONAL ARRANGEMENTS

Most available MADS analyses are located at NCAR, NCDC, and ECMWF. For example, NMC analyses and observations on magnetic tape are sent to both NCAR and NCDC within about 2 weeks of creation. In general, data sets should be stored in at least two archives in the United States for backup

---

[1] For more information about data requirements and data management strategy, refer to Jenne (1982) and CES (1989).

protection. Such data backup has already proven valuable. Having data in two archives usually adds very little to the overall cost because the data are often needed in a second working archive for reasons independent of backup considerations.

As substantial periods of model assimilation are completed, the data sets should also be available in at least two permanent archives. In addition, these analyses should be placed on "publication media," such as digital audio tapes and CD-ROM (compact disk, read-only memory) disks, so they can be sent in bulk to a number of major users at reasonable cost.

Most major U.S. archives are on magnetic tape because the media costs are less than for optical disks. However, mass storage devices are preferable so that users do not have to search tapes for specific files.

A mass storage device permits a user to place named data sets (or files) on the storage device without having to be concerned about physical location, thereby greatly simplifying data handling for the user and permitting efficient use of the device. Mass storage devices for supercomputers typically cost $2 million to $4 million. Mass storage capability in the price range of $5,000 to $150,000 could possibly be developed for limited applications by local users.

Every 6 to 12 years it becomes necessary or desirable to put existing tape-stored data sets onto new storage media. A very large benefit of proper mass storage design is that data sets can be automatically transferred to new storage media without extensive user involvement. Otherwise, this would require great effort.

FUTURE DIRECTIONS

In the future, changes in technology and cost will make it practical to store considerable amounts of data at the scientist's computer workstation, which will have the power of a 1980 Cray system. The major data archive centers will store the data and put major amounts of it on various storage media (DAT [data acquisition tape], CD-ROM, and so on). The increasing bandwidth of communications also will permit more data to be sent by electronic transfers over the High Performance Computer and Communications program links.

Access to Model Data and Observations

The modes of access should be as follows:

- Operational assimilated data sets (including delayed mode) will be transmitted routinely by operational centers to NCDC, NCAR, NODC, or other designated national archive centers.

- Data may be obtained on request from NCDC, NCAR, NODC, or other major archive centers.
- Universities may request a selection of current data transmitted in real time, via UNIDATA (University Data Broadcast Project), NOAA, or other national computer network.
- Large amounts of data should be stored at the scientist's site using DAT tapes, CD-ROM disks, or related inexpensive technology for immediate use.
- A scientist may choose to access large amounts of data directly from archival centers through computer networking.
- Archival centers should prepare and publish service-oriented mass storage media (e.g., CD-ROM disks) at cost for publishing. Standardized data access (unpacking), manipulation, and display software should be automatically included in published media and electronic transfers.
- A scientist may request selected data sets on inexpensive published mass storage media from archival centers.

Locate More Data at the Scientist's Site

Two new storage devices, CD-ROM and DAT or related technology, will permit scientists to store significant amounts of data in their own computer systems. The purchase costs of both CD-ROM and DAT hardware are expected to decrease to $50 to $200 within a few years.

Consider the time needed to prepare software for a CD-ROM or DAT. Suppose there are about 10 data sets on a CD-ROM. The time to formulate indexes to the data and prepare basic access software is about 6 weeks. A rather extensive set of display capabilities may take as long as 6 months to develop. The point is that this effort should be done once, routinely, in a nationally focused archive system, to meet the broad national needs for these data. Standard data access tools and metadata, manipulation, and imaging software should be included routinely with the assimilated data sets.

Software for Data Display and Manipulation

Many forecast centers, scientific groups, and government activities have software (e.g., NOAA/PROFS, McIDAS [Man-Computer Interactive Data Access System], UNIDATA [University Data Broadcast Project]) to view analyzed fields and observations. A basic software capability for display should be distributed along with the data. For example, if a depiction of geopotential height, temperature, or wind for a selected portion of the world is needed at a major forecast center, the computer at the center can be commanded to prepare a visual display from the assimilated data sets. Scientists have found these displays to be essential for research purposes. The key parts of the display software should be made portable so that they can be easily run on any computer workstation. Local display hardware could include, for example, multicolor map plotters, the computer monitor, or equipment for transferring displays to slides.

Funding agencies should support proposals for developing compilations of existing routines and preparing interfaces for existing data sets. Such an activity would be cost effective because it would take advantage of software already in existence.

National Weather Service Modernization Program

During the 1990s the National Weather Service (NWS) will be installing a wide range of new observational systems, including the WSR-88D Doppler radar, profiler, Automated Surface Observing System (ASOS), and advanced satellite systems. The resulting 100-fold increase in the amount of data that must be integrated into a dynamically and kinematically consistent data set will require the use of model-based assimilation systems for effective understanding and utilization of this greatly enhanced data flow.

Data assimilation systems that will utilize the new data streams of the 1990s are presently under development. The research community outside the operational centers thus has an opportunity to contribute to the design of the system and to ensure that the needs of the broad community are considered.

Advances in Computer Technology

Advances in computing technology will affect the process by which model-assimilated data sets are generated as well as the means by which they are interrogated by scientists. Greater computing power will allow more sophisticated assimilation techniques to be used at national laboratories and forecast centers and will permit modeling and measurements groups to utilize specialized data assimilation software tailored to their own needs. Routine assimilation of large earth-atmospheric-oceanic-biogeochemical data sets will become feasible as the High Performance Computer and Communications Initiative program is carried out.

Use of these data sets by scientists will be facilitated greatly by the increasing speed and storage of individual computer systems at prices affordable to individuals and small laboratories. However, facilities for local processing and graphical evaluation of large assimilated data sets will only be useful if appropriately designed software for data management is also available and ready access to data at national archives is provided.

7

Conclusions and Recommendations

The panel has reached four major conclusions. These conclusions are stated below, followed by a discussion setting forth the basis for each conclusion.

CONCLUSIONS

1. **Four-dimensional (space and time) data assimilation as a subdiscipline of geophysical sciences is fundamental for the synthesis of diverse, temporally inconsistent, and spatially incomplete observations into a coherent representation of an evolving geophysical system.**

Within the past decade, data assimilation has emerged as a robust strategy whereby prior information and current information are combined to describe a geophysical system. Temporal and spatial consistency enters in the description through a numerical assimilating model that possesses predictive, quality control, and validation capabilities, based on the governing equations for the given geophysical systems. Data assimilation produces data sets that, by virtue of their integrity and consistency, provide information and value that significantly exceed those of the incoming observations. The method provides for systematic validation of observations, synthesized data fields, and the prediction model itself.

2. **Viewed as an integral element of the scientific process, four-dimensional model assimilation of geophysical data is a systematic, quantitative, objective, iterative means of inference and testing aimed at advancing understanding and prediction of nonlinear dynamical geophysical systems where interactions occur continually among relevant physical, chemical, and biogeochemical processes.**

Use of data assimilation will advance understanding of geophysical phenomena governed by nonlinear, time-dependent dynamics as opposed to time-invariant conditions. Through systematic confrontation between observations and a priori understanding, as expressed by nonlinear governing equations and conceptual models, data assimilation provides a unique learning tool for promoting scientific understanding of the earth system. The learning process involved is inherently open ended and iterative.

Current developments in data assimilation methodology will permit the use, continuously in time, of data that are nonlinearly related to model variables. Data assimilation is the only objective way to synthesize four-dimensional, multifaceted data that describe atmospheric, oceanic, and land surface processes. Systematic and iterative confrontation of model-predicted states on all time scales with observations (the learning process) has exposed and will continue to expose defects in both the data and the models. This confrontation leads to model improvement by identifying missing processes and critical observation gaps, by providing estimates of the magnitude of individual processes, and by providing estimates of the feedbacks between processes. The learning process leads to improvements in observational systems and sampling strategies.

Within this iterative process, the evaluation of assimilated data by research users, including insertion of limitations, shortcomings, and critical comments into the archive record of the data sets, is important. Such action should become a routine step in studies in order to stem the propagation of error in subsequent analysis should error be established.

3. **Since the physical and dynamical consistency of model-assimilated data sets results in a level of information and added value that significantly exceeds that of the incoming observations, assimilation data sets have been and will be used even more extensively by the scientific community for diagnostic, predictive, and process studies, supplemented by original observations when needed.**

A major part of the progress achieved in operational weather prediction in recent years must be credited to improvements in data assimilation and the increasing use of model-assimilated and simulated data sets in research.

For example, the data sets produced by the National Oceanic and Atmospheric Administration's (NOAA) Geophysical Fluid Dynamics Laboratory (GFDL) and the European Centre for Medium Range Weather Forecasts (ECMWF), along with the operational global, hemispheric, and regional data sets produced by the National Meteorological Center (NMC), have been the principal basis for improvements in forecast skill, weather prediction research, studies of climatic processes, diagnostic studies of the planetary circulation, and interacting, coupled atmospheric-oceanic models. Studies of mesoscale circulation systems in the atmosphere and oceans are increasingly using both assimilated and simulated data sets produced by high-resolution models.

Model-assimilated data sets are internally consistent with the dynamics and physical processes of the atmosphere as represented by the models employed at the operational and research centers. These data sets use all available observations to provide the best possible description of the atmosphere in a compact, orderly, gridded, or spectral format. As such, they form an exceedingly valuable resource that has become the data set of choice for the great majority of atmospheric researchers. Model-assimilated data sets for the ocean, land surface, and other components of climate change and global change processes should prove to be equally valuable for research once they have been developed.

4. **Atmospheric model data assimilation is a proven strategy in an advanced state of development. Although still in an early stage of development for the earth sciences as a whole, model data assimilation is strategically situated to address the pressing national need to observe and understand global change as it occurs in the coming decades. The immediate need is to develop and provide a nationally focused capability to synthesize, test, and utilize for understanding and prediction the information from the greatly enhanced new ground-based and satellite observing systems already scheduled for the next two decades.**

Implementation of advanced observation technology for the earth sciences is rapidly surpassing the capability of the scientific community to manage and utilize effectively the potentially overwhelming output of global data in the coming decades. NOAA, for example, is rapidly implementing, as part of the National Weather Service (NWS) Modernization Program, new networks of WSR-88D Doppler radars, atmospheric profilers, and advanced satellites that together will increase the flow of meteorological and oceanic data by up to 100 times. The National Aeronautics and Space Administration's (NASA) Earth Observing System (EOS), together with the Upper Atmosphere Research Satellite (UARS), will provide observations at a rate and volume never before seen. These data streams must be managed through a nationally focused program that produces model-assimilated data sets that provide a high-quality, valuable, coherent, and integrated understanding of the earth system as a whole.

RECOMMENDATIONS

The panel recommends the following strategy for a nationally focused program for four-dimensional model assimilation of data for the earth system.

1. **The strategy for the nationally focused program will:s**

- **Develop and expand applications of model-based data assimilation efforts to interdisciplinary areas that are necessary for integrated earth-ocean-atmosphere-biogeochemical models.**
- **Implement an integrated, multicenter, nationally focused archive system for model-assimilated and-tested data sets and provide for ready access to these sets by the scientific community over the long term.**
- **Provide for continuing scientific exchange and collaboration among the various groups and individuals engaged in geophysical data assimilation, particularly scientists involved with the Earth Observing System Data and Information System (EOSDIS); the national meteorological, oceanographic, and climate centers; and the High Performance Computer and Communications Initiative.**
- **Provide for the generation of routine research-quality, model-assimilated and-tested geophysical data sets to serve a broad range of national endeavors, including climate and global change research and predictions.**
- **Establish a working group to develop an implementation plan for this nationally focused program.**

Given the growing national focus on societal problems stemming from climate and global change, and the greatly enhanced observational capabilities already planned and scheduled for the 1990s and beyond, the need exists nationally to utilize global geophysical information to address these problems. This need to detect and monitor climate and global change as it occurs places unusual demands on both observing and data assimilation capabilities. Development and application of global and regionally assimilated data sets prepared by state-of-the-art prediction models for understanding the earth-ocean-atmosphere system and the physical processes that determine its evolution require a nationally focused program. To date, the production of gridded analyses using data assimilation has largely been driven by numerical weather prediction activities located within operational weather centers. The production of optimum research-quality assimilated sets requires a delayed mode of analysis whereby all available relevant geophysical data are utilized. Within this larger effort, use of the quality control and validation information gained from operational assimilation is essential as a two-step process for production of optimum global data sets. Such efforts must be embodied in a nationally focused program to meet the varied needs of academic, government, and public sectors concerned with climate and global change, in making important economic and political policy decisions to cope with the rate of change predicted. Model data assimilation applied to these issues would provide the capability to synthesize these heterogeneous data consistently and objectively. A nationally focused program for data assimilation would ensure an archiving system that would make these coherent data sets readily available for the broad range of national needs.

The strategy and implementation plan should also include the following specific actions.

2. **To validate and maintain quality control of new types of remotely sensed and experimental in situ geophysical data, including research data from field programs, the application of operational data assimilation models is essential. Funding agencies should routinely provide sufficient funds and computer capacity for this purpose, including provision for timely communication of such new and experimental data sets to designated operational assimilation centers.**

Over the past two decades, operational data assimilation for global weather forecasting has been developed to the point where it is invaluable for quality control and validation of remotely sensed and experimental data. The method ensures an internally consistent synthesis of all available data and includes the means for validating observational data streams against other data types. These capabilities have been systematically exploited to identify errors and biases in many types of in situ and remotely sensed data. Quality control and validation information from operational assimilation are important for producing research-quality data sets.

Validations, both global and regional, performed in this way consider a broader range of atmospheric conditions than traditional field-validation campaigns. The ready provision of new research data to operational centers can provide for the timely preparation of assimilated data sets, including immediate quality control and consistency checks, the timely nature of which is important for the conduct of experiments. As such, this new approach to validation and quality monitoring is a valuable supplement to field campaigns. These benefits can only be realized, however, if the research data are transmitted to suitable operational centers in a timely fashion and assimilated data sets are likewise returned.

3. **A coordinated national program should be implemented and funded to develop consistent, long-term assimilated data sets (extended back to about 1950) for the study of climate and global change. This effort will reanalyze with a state-of-the-art global- and regional-scale data assimilation model all atmospheric and oceanic data available since about 1950 in order to produce the best possible, validated, temporally and spatially consistent data sets for the study of climate and global change. Model biases should be identified and eliminated as far as possible at this time. This reanalysis should be repeated as advances in the state of the art require.**

Existing global model-assimilated data sets for climate studies are compromised by changes in assimilating models, in methods of dealing with raw data, in the observing networks and systems, and in assimilation procedures. These factors introduce discontinuities and inconsistencies in long time series of model-assimilated geophysical data, originally produced for weather prediction purposes, that make them only marginally useful for climate and global change analysis and predictions.

An analysis of the entire useful global climate record from about 1950, using a single state-of-the-art data assimilation system, is the optimum means for producing temporally and spatially consistent and continuous global model-assimilated data sets useful for climate and global change purposes. Since data sets of this quality do not currently exist, a comprehensive reanalysis of the existing observed and analyzed data is needed now. During this effort, changes in observing capabilities (such as the inclusion of global satellite data beginning in the 1970s) can be more accurately assimilated and the results reflected in statistics of climate and global change. The lack of satellite data globally in the 1950s and early 1960s, as well as other changes in observing capabilities, introduces some degree of uncertainty in the existing data sets. As assimilation strategies and computers are developed in the future, further reanalyses with more advanced data assimilation models will be more useful. Future reanalyses will provide an orderly way to maintain consistency and eliminate model biases, thus increasing the reliability of the climatic record.

The model used in any extended analysis or reanalysis of the record should be the best available at the time of analysis. Currently, such an analysis could probably be conducted using an operational or advanced research global atmospheric general circulation model (GCM) with sea surface temperature (SST) specified by the observed values. Future analyses will require increasingly sophisticated models with more processes included, eventually culminating in a fully coupled atmosphere-ocean-land-biosphere-cryosphere model that assimilates all forms of data (National Research Council, 1990).

To make these long climatic analyses possible, the observed data must be collected and archived in uniform and accessible formats, with comprehensive documentation of instrumental calibration and other characteristics that affect data quality. Because each successive analysis would be of the entire climate record to that date, special efforts to archive observed data in a convenient, compact electronic format will be needed.

4. **Data assimilation models for the mesoscale should be developed in concert with regional prediction models. The systems should integrate data from the enhanced observational capabilities of the 1990s that will be provided through the NWS Modernization Program, advanced satellite systems, and the EOS. These models should be capable of integrating specially observed data with the conventional data stream and also provide for realistic responses to various types of system forcing for the time and space scales emphasized. The mesoscale assimilation systems should also ensure effective nesting with larger-scale analysis systems.**

Use of model-based assimilation systems for the generation of mesoscale data sets is especially appropriate because many new mesoscale data sources can resolve smaller spatial and temporal scales than current synoptic-scale operational models are designed to do.

High-resolution atmospheric assimilation systems have the valuable ability to generate mesoscale structures that are forced by processes such as differential surface heat, moisture and momentum fluxes, topographic variations, and latent and radiative heating. The assimilating model should include complete physics in order to simulate mesoscale structures and processes that are not directly observed. Since the lateral boundaries of an analyzed mesoscale domain must frequently be shifted in time to follow the evolution of mesoscale disturbances, the mesoscale assimilation system must be effectively and accurately linked with a coarser-scale (e.g., global and synoptic) model assimilation system.

5. **In order to provide ready access by the scientific community, a multicenter geophysical data archive system, electronically linked for maximum effectiveness for assimilated dataset management, transfer, and usage, should be created.**

- **The data management activity for the multicenter archiving system will ensure the routine compilation and availability of**

observed, model-assimilated, and model-predicted data sets in a structural format jointly adopted and coordinated with the EOSDIS. There is a vital need to archive together with the data sets the modeling codes, information on the input data and how they were processed, and so forth for the scientist to be able to evaluate and understand the archived data sets.

- **Accessibility to archived model-assimilated data sets should be part of a preplanned, service-oriented archive system, including on-line electronic links, low-cost publication media for use on individual workstations, routine inclusion of metadata and software for unpacking and manipulating data sets, and high-quality imaging capabilities.**

The assembly of a properly formatted and sequenced interdisciplinary assimilated data archive that will span approximately the last 40 years and be extended into the future is a major and crucial task for studies of climate and global change and will require a designated, funded management activity. The archive should consist of data used for operational purposes (e.g., numerical weather prediction) as well as data received in a delayed mode from a variety of national and international sources. The archive should also include model-simulated data that are to be used in scientific and validation studies involving intercomparisons with observed and assimilated data. Assembly of such an archive from a multitude of different data collection points is relatively difficult and time consuming. Therefore, insofar as feasible with respect to efficiency and detail, most data should be collected through the operational system in real time. Delayed data also need to be gathered systematically before they are scattered among too many sub-archives. The goal should be to prepare comprehensive global data sets, with each observational component in a good format and in the sequential order required for long-term analyses. This activity needs to accommodate the diversity of data types, which include rawinsonde data, ship marine data, aircraft reports, cloud drift winds, drifting buoys, expendable bathythermographs (XBTs), satellite retrieval information on atmospheric and sea surface conditions, sea ice coverage, river discharge, and land surface conditions.

Data sets should be readily available to users on request from archives. Advanced methods are needed to lower the cost of obtaining data and to facilitate the user's access to the data. Considerable portions of the archived data sets should be "published" on media devices such as data acquisition tapes (DATS) or compact disks that can be readily reproduced at low cost. Software routines that access data should accompany the device, and selected browse-and-display routines also should be available. A display option is a converter that costs less than $10,000 and has the ability to prepare video cassette recorder tapes.

6. **Interdisciplinary and discipline graduate education and research opportunities in four-dimensional geophysical data assimilation should be created. To provide essential expertise to cope with the 100-fold increases in observational data from greatly enhanced observing systems in the 1990s and beyond, the development of graduate education, including advanced degree and research opportunities in the fundamental subdiscipline of four-dimensional data assimilation, is needed. Immediate support should be provided for graduate courses and research fellowships in university departments of atmospheric, oceanic, global change, and related sciences. These programs should be explicitly coordinated with related system development, such as EOSDIS, the High Performance Computer and Communications program, and national computer-linked networks.**

A serious shortage of research scientists with the background, interest, and ability to develop advanced four-dimensional data assimilation methodology and use it in producing high-quality model-assimilated data sets exists at present. The 10^2 to 10^3 increase in global data in the 1990s will require at least a 10- to 100-fold increase in data assimilation modeling, the development and utilization of which will require an equal increase in the number of scientists engaged in these efforts. Graduate students and junior researchers need to be made aware of the problems and challenges of four-dimensional data assimilation for the earth science system and be provided with opportunities to develop their abilities in this area. The central importance of data assimilation in advancing understanding in the atmospheric, oceanic, and related sciences must be stated and endorsed through support for graduate courses and research fellowships at universities in order to attract students and young scientists to this crucial effort that seeks to understand climate and global change.

References

Anderson, D.L.T., and A.M. Moore. 1989. Initialization of equatorial waves in ocean models. J. Phys. Oceanogr. 19:116–121.

Anderson, D., A. Hollingsworth, S. Uppala, and P. Woiceshyn. 1991. A study of the use of scatterometer data in the European Centre for Medium Range Weather Forecasts operational analysis-forecast model, 1, Quality assurance and validation. J. Geophys. Res. 96:2619–2634.

Andersson E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux, and Z. Zhang. 1991. Observing system experiments on NESDIS statistical retrievals of TOVS satellite data using the 1988 ECMWF data assimilation system. To appear in Mon. Weather Rev., July issue.

Andre, J.-C., J.-P. Goutorbe, and A. Perrier. 1986. HAPEX-MOBILHY: A hydrologic atmospheric experiment for the study of water budget and evaporation flux at the climatic scale. Bull. Am. Meteorol. Soc. 67:138–145.

Arkin, P.A., and P.E. Ardanuy. 1989. Estimating climatic-scale precipitation from space: A review. J. Climate 2:1229–1238.

Arpe, K. 1991. The hydrological cycle in the ECMWF short range forecasts. Accepted by Dyn. Atmos. Oceans.

Atlas, R., A.J. Busalacchi, M. Ghil, S. Bloom, and E. Kalnay. 1987. Global surface wind and flux fields from model assimilation. J. Geophys. Res. 92:6477–6487.

Bennett, A.F., and W.P. Budgell. 1987. Ocean data assimilation and the Kalman filter, spatial regularity. J. Phys. Oceanogr. 17:1583–1601.

Bennett, A.F., and P.C. McIntosh. 1982. Open ocean modeling as an inverse problem: Tidal theory. J. Phys. Oceanogr. 12:1004–1018.

Bierman, G.J. 1977. Factorization Methods for Discrete Sequential Estimation. Academic Press. Orlando, Fla., 241 pp.

Bucy, R.S., and P.D. Joseph. 1987. Filtering for Stochastic Processes with Applications to Guidance, 2nd ed. Chelsea, New York, 217 pp.

Budgell, N.P. 1986. Nonlinear data assimilation for shallow water equations in branched channels. J. Geophys. Res. 91:10633–10644.

Cane, M.A., S.E. Zebiak, and S.C. Dolan. 1986. Experimental forecasts of El Niño. Nature 321:827–832.

CES. 1989. Our Changing Planet: The FY 1990 Research Plan. A report by the Committee on Earth Sciences. Federal Coordinating Council on Science, Engineering, and Technology, Office of Science and Technology Policy, Washington, D.C., 181 pp.

Charney, J.G., M. Halem, and R. Jastrow. 1969. Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci. 26:1160–1163.

Davies, H.C., and R.E. Turner. 1977. Updating prediction models by dynamical relaxation: An examination of the technique. Q. J. R. Meteorol. Soc. 103:225–245.

Derber, J.C. 1989. A variational continuous assimilation technique. Mon. Weather Rev. 117:2437–2446.

Derber, J., and A. Rosati. 1989. A global oceanic data assimilation system. J. Phys. Oceanogr. 19:1333–1347.

Gandin, L.S. 1963. Objective Analysis of Meteorological Fields. Leningrad, Gidrometeoizdat, in Russian. (English Translation: Israel Program for Scientific Translations, Jerusalem, 1965, 242 pp.)

Gelb, A., ed. 1974. Applied Optimal Estimation. M.I.T. Press, Cambridge, Mass., 374 pp.

Ghil, M. 1989. Meteorological data assimilation for oceanographers. Part I: Description and theoretical framework. Dyn. Atmos. Oceans 13:171–218.

Ghil, M., and S. Childress. 1987. Topics in Geophysical Fluid Dynamics, Atmospheric Dynamics, Dynamo Theory and Climate Dynamics. Springer-Verlag, New York, 485 pp.

Ghil, M., and P. Malanotte-Rizzoli. 1991. Data assimilation in meteorology and oceanography. Adv. Geophys., in press.

Ghil, M., M. Halem, and R. Atlas. 1979. Time-continuous assimilation of remotesounding data and its effect on weather forecasting. Mon. Weather Rev. 107:140–171.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson. 1981. Applications of estimation theory to numerical weather predictions. In Dynamic Meteorology: Data Assimilation Methods, L. Bengtsson, M. Ghil, and E. Kallen, eds. Springer-Verlag, New York, pp. 139–224.

Golding, B.W. 1989. The Meteorological Office experimental mesoscale numerical weather prediction system: July 1989. Technical Note No. 30, British Meteorological Office, Bracknell, U.K. 26 pp.

Haidvogel, D.G., and A.R. Robinson, eds. 1989. Spatial Issue on Data Assimilation in Ocean Modeling. Dyn. Atmos. Oceans 13(3 and 4):171–575.

Harrison, D.E., W.S. Kessler, and B.J. Giese. 1989. Ocean circulation model hindcasts of the 1982–83 El Niño: Thermal variability along the ship-of-opportunity tracks . J. Phys. Oceanogr. 19:397–418.

Hoffman, R.N., and E. Kalnay. 1983. Lagged-average forecasting, an alternative to Monte Carlo forecasting. Tellus 35A:100–118.

Hoke, J.E., and R.A. Anthes. 1976. The initialization of numerical models by a dynamic initialization technique. Mon. Weather Rev. 104:1551–1556.

Hoke, J.E., N.A. Phillips, G.J. DiMego. J.J. Tuccillo, and J.G. Sela. 1989. The regional analysis and forecast system of the National Meteorological Center. Weather and Forecasting 4:323–334.

Hollingsworth, A. 1987. Objective analysis for numerical weather prediction. Proceedings of International Symposium for Short-and Medium-Range Weather Prediction, WMO/IUGG NWP symposium, Tokyo. August 4–8, 1986, T. Matsuno, ed. J. Meteorol. Soc. Jap:1–60.

Hollingsworth, A. 1989. The role of four-dimensional assimilation in the quality control, interpretation, and synthesis of climate data. Oceanic Circulation Models: Combining Data and Dynamics, J. Willebrands and D. Anderson, eds. Kluwer Academic Publishers, pp. 303–343 .

Hollingsworth, A., and P. Lönnberg. 1986. The statistical structure of short range forecast errors as determined from radiosonde data. Part I. The wind field. Tellus 38A:111–136.

Hollingsworth, A., and P. Lönnberg. 1989. Verification of objective analyses: Diagnosis of analysis system performance. Meteorol. Atmos. Phys. 40:3–27.

Hollingsworth, A., A.C. Lorenc, M.S. Tracton, K. Arpe, G. Cats, S. Uppala, and P. Kallberg. 1985. The response of numerical weather prediction systems to FGGE level IIb data. Part I. Analyses. Quart. J. Roy. Meteorol. Soc. 111:1–66.

Hollingsworth, A., D.B. Shaw, P. Lönnberg, L. Illari, K. Arpe, and A.J. Simmons. 1986. Monitoring of observation and analysis quality by a data assimilation system. Mon. Weather Rev. 114.861–879.

Horel, J.D., and J.M. Wallace. 1981. Planetary scale phenomena associated with the southern oscillation. Mon. Weather Rev. 109:813–829.

Illari, L. 1989. The quality of satellite precipitable water content data and their impact on analyzed moisture fields. Tellus 41A:319–337.

Janssen, P.A.E.M., P. Lionello, M. Reistad, and A. Hollingsworth. 1989. Hindcasts and data assimilation studies with the WAM model during the SEASAT period. J. Geophys. Res. 94:973–993.

Jarwinski, A.H. 1970. Stochastic Processes and Filtering Theory. Academic Press, New York, 376 pp.

Jenne, R.L. 1982. Planning Guidance for the World Climate Data System. WCP-19, World Meteorological Organization, Geneva, 136 pp.

Julian, P. 1984. Objective analysis in the tropics: A proposed scheme. Mon. Weather Rev. 112:1752–1767.

Kalman, R.E. 1960. A new approach to linear filtering and prediction problems. Trans. ASME J. Basic Eng. 82D:35–45.

Kalman, R.E., and R.S. Bucy. 1961. New results in linear filtering and prediction theory. Trans. ASME J. Basic Eng. 83D:95–108.

Kalnay, E., R. Atlas, W. Baker, and J. Susskind. 1985. GLAS experiments on the impact of FGGE satellite data on numerical weather prediction. Proceedings of the First National Workshop on the Global Weather Experiment: Current Achievements and Future Directions, Vol. 2, Part I, pp. 121–145. National Academy Press, Washington, D.C.

Kasahara, A., R.C. Balgovind, and B.B. Katz. 1988. Use of satellite radiometric imagery data for improvement in the analysis of divergent wind in the tropics. Mon. Weather Rev. 116:866–883.

Kelly, G., E. Andersson, A. Hollingsworth, P. Lönnberg, J. Pailleux, Z. Zhang. 1991. Quality control of NESDIS physical retrievals of TOVS satellite data. To appear in Mon. Weather Rev., July issue.

Krishnamurti, T.N., and S. Low-Nam. 1986. On the relationship between the outgoing longwave radiation and the divergent circulation. J. Meteor. Soc. Jpn. 64:709–719.

Lange, A.A. 1989. Results of the WMO/CAS NWP Data Study and Intercomparison Project for forecasts for the northern hemisphere in 1988. World Weather Watch Technical Report No. 7, WMO/TD No. 317. World Meteorological Organization, Geneva.

Lazier, J.R. 1988. Temperature and salinity changes in the deep Labrador Sea, 1962–1986. Deep-Sea Res. 35:1247–1253.

Le Dimet, F.X., and O. Talagrand. 1986. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. Tellus 38A:97–110.

Leetmaa, A., and M. Ji. 1989. Operational hindcasting of the tropical Pacific. Dyn. Atmos. Oceans 13:465–490.

Leith, C.E. 1974. Theoretical skill of Monte Carlo forecasts. Mon. Weather Rev. 102:409–418.

Levitus, S. 1989. Interpentadal variability of temperature and salinity at intermediate depths of the North Atlantic Ocean. J. Geophys. Res. 94:6091–6131.

Lewis, J.M., and J.C. Derber. 1985. The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus 37A:309–322.

Long, R.B., and W.C. Thacker. 1989a. Data assimilation into a numerical equatorial ocean model. Part I: The model and assimilation algorithm. Dyn. Atmos. Ocean. 13:379–412.

Long, R.B., and W.C. Thacker. 1989b. Data assimilation into a numerical equatorial ocean model. Part II: Assimilation experiments. Dyn. Atmos. Ocean. 13:413–440.

Lorenc, A.C. 1976. Results of some experiments assimilating observations from a simulation of the FGGE observing system into a global general circulation model. In JOC Study Conference on Four-Dimensional Data Assimilation, Paris, 1975. Report II, GARP on Numerical Experimentation. Pub. WMO/ICSU, pp. 358–374. World Meteorological Organization, Geneva.

Lorenc, A.C. 1981. A global three-dimensional multivariate statistical interpolation scheme. Mon. Weather Rev. 109:701–721.

Lorenc, A.C. 1986. Analysis methods for numerical weather prediction. Q. J. R. Meteorol. Soc. 112:1177–1194.

Lyne, W.H., R. Swinbank, and N.T. Birch. 1982. A data assimilation experiment and the global circulation during the FGGE Special Observing Periods. Q. J. R. Meteorol. Soc. 108:575–594.

Marchuk, G.I. 1974. Numerical solution of the problems of the dynamics of the atmosphere and ocean. Leningrad, Gidrometeoizdat, 303 pp. (in Russian).

McPherson, R.D., K.H. Bergman, R.E. Kistler, G.E. Rasch, and D.S. Gordon. 1979. The NMC operational global data assimilation system. Mon. Weather Rev. 107:1455–1461.

Meteorological Office. 1987. The storm of 15/16 October 1987. Meteorological Office Report, December 1987. Available from Her Majesty's Stationery Office, London, ISBN-0-86180-232-2.

Miller, R.N. 1986. Toward the application of the Kalman filter to regional open ocean modeling. J. Phys. Oceanogr. 16:72–86.

Miller, R.N., and M.A. Cane. 1989. A Kalman filter analysis of sea level height in the tropical Pacific. J. Phys. Oceanogr. 19:773–790.

Miyakoda, K., L. Umscheid, D.H. Lee, J. Sirutis, R. Lusen, and F. Pratte. 1976. The near-real time global four-dimensional analysis experiment during the GATE period, Part I. J. Atmos. Sci. 33:561–571.

Moore, A.M. 1989. Aspects of geostrophic adjustment during tropical ocean data assimilation. J. Phys. Oceanogr. 19:435–461.

Moore, A.M., N.S. Cooper. and D.L.T. Anderson. 1987. Initialization and data assimilation in models of the Indian Ocean. J. Phys. Oceanogr. 17:1965–1977.

Morlire, A.P., G. Reverdin, and J. Merle. 1989. Assimilation of temperature profiles in a general circulation model of the tropical Atlantic. J. Phys. Oceanogr. 19:1892–1899.

National Research Council, Committee on Global Change. 1990. Research Strategies for the U.S. Global Change Research Program. National Academy Press, Washington, D.C., 291 pp.

National Research Council, Panel on Dynamic Extended Range Forecasting. 1991. Prospects for Extending the Range of Prediction of the Global Atmosphere. National Academy Press, Washington, D.C., in press.

Palmer, T.N. 1986. Influence of Atlantic, Pacific, and Indian oceans on Sahel rainfall. Nature 322:251–253.

Parrish, D.F., and S.E. Cohn. 1985. A Kalman Filter for a Two-Dimensional Shallow-Water Model: Formulation and Preliminary Experiments. Office Note 304, National Meteorological Center, Washington, D.C., 64 pp.

Penenko, V., and N.N. Obraztsov. 1976. A variational initialization method for the fields of the meteorological elements. Meteorologiya i Gidrologiya (English translation) 11:1–11.

Philander, S.G.H., W.J. Hurlin, and R.C. Pacanowski. 1987. Initial conditions for a general circulation model of the tropical ocean. J. Phys. Oceanogr. 17:147–157.

Rasmusson, E.M. 1968. Atmospheric water vapor transport and the water balance of North America. II. Large-scale water balance investigations. Mon. Weather Rev. 96:720–734.

Reynolds, R.W., K. Arpe, C. Gordon, S.P. Hayes, A. Leetmaa, and M.J. McPhaden. 1989. A comparison of tropical Pacific surface wind analyses. J Climate 2:105–111.

Robinson, A.R., and W.G. Leslie. 1985. Estimation and prediction of oceanic eddy fields. Prog. Oceanogr. 14:485–510.

Robinson, A.R., J.A. Carton, N. Pinardi, and C.N.K. Mooers. 1986. Dynamical forecasting and dynamical interpolation: An experiment in the California current. J. Phys. Oceanogr. 16:1561–1579.

Robinson, A.R., M.A. Spall, L.J. Walstead, and W.G. Leslie. 1989. Data assimilation and dynamical interpolation in gulfcast experiments. Dyn. Atmos. Oceans 13:301–316.

Seaman, N.L. 1989. Description and evaluation of the meteorological cases used in the RADM NAPAP evaluation and assessment. Final Report to the Environmental Protection Agency under Interagency Agreement DW49933202, Pennsylvania State University, 245 pp.

Sellers, P.J., F.G. Hall, G. Asar, D.E. Strebel, and R.E. Murphy. 1988. The First ISLSCP Field Experiment (FIFE). Bull. Am. Meteorol. Soc. 69:22–27.

Shukla, J. 1985. Predictability. In Advances in Geophysics, 28 (Issues in Atmospheric and Oceanic Modeling. Part B), S. Manabe, ed. Academic Press, New York, pp. 87–122.

Stern. W.F., and J.J. Ploshay. 1991. A scheme for continuous data assimilation. Mon. Weather Rev., in press.

Stern, W.F., J.J. Ploshay, and K. Miyakoda. 1985. Continuous data assimilation at GFDL during FGGE. In Report of Seminar/Workshop on Data Systems and Observing System Experiments with Particular Emphasis on FGGE, in 1984, sponsored by ECMWF and WMO, Vol. 2, pp. 125–156. European Centre for Medium Range Weather Forecasts, Bracknell, U.K.

Talagrand, O., and P. Courtier. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I. Theory. Q. J. R. Meteorol. Soc. 113:1311–1328.

Thacker, W.C., and R.B. Long. 1988. Fitting dynamics to data. J. Geophys. Res. 93:1227–1240.

Tiedtke, M., W.A. Heckley, and J. Slingo. 1988. Tropical forecasting at ECMWF. Q. J. R. Meteorol. Soc. 144:639–664.

Tropical Rainfall Measuring Mission Science Steering Group. 1988. Report of the Science Steering Group for a Tropical Rainfall Measuring Mission (TRMM). J. Simpson, ed., NASA, Washington, D.C., 94 pp.

Uppala, S., P. Kalberg, A. Hollingsworth, and S. Tibaldi. 1985. Results from two recent observing system experiments. Proceedings of the First National Workshop on the Global Weather Experiment: Current Achievements and Future Directions, Vol. 2. Part I. pp. 85–117. National Academy Press, Washington, D.C.

Wallace, J.M., and D.S. Gutzler. 1981. Teleconnections in the geopotential height field during the northern hemisphere winter. Mon. Weather Rev. 109:785–812.

World Climate Research Program. 1986. Report of the Workshop on Global Large-Scale Precipitation Data Sets for the World Climate Research Programme. WCP-111, WMO/TD-No. 94, World Meteorological Organization, Geneva, 50 pp.

List of Acronyms

AIRS	Atmospheric Infrared Sounder
ASOS	Automated Surface Observing System
CAC	Climate Analysis Center, NOAA National Meteorological Center
CD-ROM	Compact Disk, Read-Only Memory
CIRES	Cooperative Institute for Research in Environmental Sciences, University of Colorado
CLIMAP	Climate: Long-Range Investigation, Mapping, and Prediction
COADS	Comprehensive Ocean-Atmosphere Data Set
DAT	Data Acquisition Tape
DERF	Dynamic Extended-Range Forecasts
ECMWF	European Centre for Medium Range Weather Forecasts
ENSO	El Niño-Southern Oscillation
EOS	Earth Observing System, NASA
EOSDIS	Earth Observing System Data and Information System
EPA	Environmental Protection Agency
ERF	Extended-Range Forecasting
ERL	Environmental Research Laboratories, NOAA
ESA	European Space Agency
FGGE	First GARP Global Experiment (see GWE)
FIFE	First ISLSCP Field Experiment

FNOC	Fleet Numerical Oceanography Center, U.S. Navy
FOCAL	Français Ocean et Climat dans l'Atlantique Equatorial
FSL	Forecast Systems Laboratory, NOAA
GARP	Global Atmospheric Research Program
GCM	General Circulation Model
GEWEX	Global Energy and Water Cycle Experiment
GFDL	Geophysical Fluid Dynamics Laboratory, NOAA
GLA	Goddard Laboratory for Atmospheres, NASA
GPCP	Global Precipitation Climatology Project
GTS	Global Telecommunications System, WWW
GWE	Global Weather Experiment (initially designated as the First GARP Global Experiment, FGGE)
HAPEX	Hydrological-Atmospheric Pilot Experiment
ISLSCP	International Satellite Land Surface Climatology Project
JGOFS	Joint Global Ocean Flux Study
LAPS	Local Analysis and Prediction System, Colorado State University
LAWS	Laser Atmospheric Wind Sounder
LFV	Low-Frequency Variability
LIMS	Limb Infrared Monitor of the Stratosphere
LINNMI	Linear Normal-Mode Initialization
LRF	Long-Range Forecasting
MADS	Model-Assimilated Data Set
MAPS	Mesoscale Analysis and Prediction System
McIDAS	Man-Computer Interactive Data Access System
MOODS	Master Ocean Observations Data Set
MRF	Medium-Range Forecasting
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency (of Japan)
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center, NOAA (World Data Center A, WMO)
NMC	National Meteorological Center, NOAA, U.S. National Weather Service
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanographic Data Center, U.S. Navy
NWP	Numerical Weather Prediction
NWS	National Weather Service
OI	Optimal Interpolation
OLR	Outgoing Long-Wave Radiation
PNA	Pacific-North America Low-Frequency Atmospheric Oscillation
RAFS	Regional Analysis and Forecast System, NWS

SAMS	Stratospheric and Mesospheric Sounder
SBUV	Solar Backscatter Ultraviolet Spectrometer
SEASAT-A	U.S. Oceanographic Satellite Monitoring System
SEQUAL	Seasonal Equatorial Atlantic Experiment
SMHI	Swedish Meteorological and Hydrological Institute
SO	Southern Oscillation
SST	Sea Surface Temperature
TOGA	Tropical Ocean and Global Atmosphere Program, WMO
TOMS	Total Ozone Mapping Spectrometer
TRMM	Tropical Rainfall Measuring Mission
UARS	Upper Atmosphere Research Satellite
UCAR	University Corporation for Atmospheric Research
UKMO	United Kingdom Meteorological Office
UNIDATA	University Data Broadcast Project
WCRP	World Climate Research Program, WMO
WMO	World Meteorological Organization, Geneva
WOCE	World Ocean Circulation Experiment
WSR-88D	U.S. National Weather Service 1988 Doppler Weather Radar System (initially designated in planning documents as NEXRAD, Next Radar System)
WWW	World Weather Watch, WMO
XBT	Expendable Bathythermograph